



2. UTGAVE OPPDATERT
OG UTVIDET

DATA DRIVEN CONSTRUCTION

NAVIGERING I DATAALDEREN I
BYGGEBRANSJEN

MED BRUKSTILFELLER FOR
KUNSTIG INTELLIGENS OG
LLM

Artem Boiko

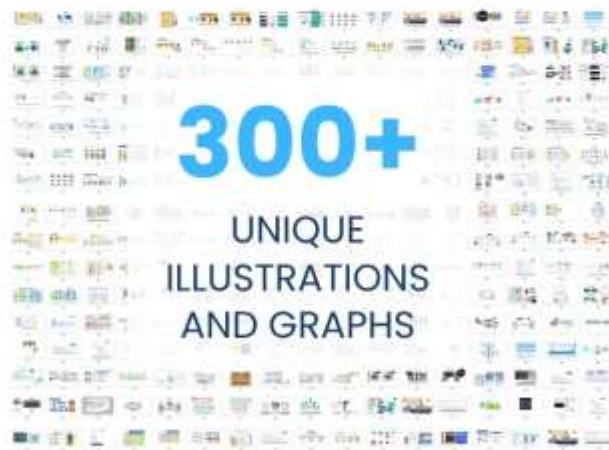


100+
KEY DATA-
RELATED
TOPICS



50+

PRACTICAL
PROBLEMS
SOLVED



300+

UNIQUE
ILLUSTRATIONS
AND GRAPHS

DATA-DRIVEN CONSTRUCTION

Navigering i dataalderen
i byggebransjen

Andre utgave, revidert og supplert

ARTEM BOIKO

“

"Boiko er IT-bransjens James Carville - i sistnevntes mye siterte "It's the economy, stupid" er det bare ett ord som trenger å byttes ut med denne berømte boken. "Det er dataene, dumme." (Og for å finne frem i datauniverset er et ordtak fra de gamle romerne, som går tilbake til gresk, fortsatt gyldig i dag: "Navigare necesse est". Forfatteren navigerer sine leserne gjennom datahavets dybder og grunner med sikker hånd og et urokkelig kompass, for ikke å snakke om en omfattende historisk tilnærming, og sist, men ikke minst, en svært original grafikk og en god sans for humor som ikke bare er åpenbar ved første øyekast. Den internasjonale responsen på Boikos bok spenner fra euforisk begeistring til temmelig galleaktig skepsis, noe som har gjort den andre tyske utgaven av boken godt. Boiko er en original og udogmatisk datatenker. Han presenterer leseren for spennende innsikter og alltid modige, til og med provoserende teser som inspirerer til videre tenkning. Utmerket medisin mot den tyske sykdommen latent konsensualisme. Det latinske ordtaket ovenfor har for øvrig et supplement: "vivere non est necesse". Det gjelder ikke for Boikos tilnærming til dataverdenen - data lever, og deres liv er nødvendig, for ikke å si avgjørende."

- Dr. Burkhard Talebitari, frilansredaktør - blant annet for tidsskriftet: BIM, som Ernst & Sohn har utgitt hvert år siden 2013.

"Artem Boikos bok er en milepæl for demokratiseringen av digitaliseringen i byggebransjen - og en virkelig "game changer" for små og mellomstore bedrifter (SMB-er). Spesielt banebrytende: Ved å bruke moderne open source-verktøy med og uten åpen kildekode kan bedrifter allerede i dag effektivt integrere data i forretningsprosessene sine og analysere dem på en lønnsom måte - helt uten inngående programmeringskunnskaper. Dette gjør den kostbare bruken av tungvinte kommersielle programverpakker overflødig. Denne boken er en oppfordring til handling! Den er en verdifull guide for alle som ikke bare ønsker å forstå den digitale transformasjonen i byggebransjen, men som også aktivt ønsker å forme den - pragmatisk, effektivt og fremtidsrettet. Nå er tiden inne for å samarbeide om å dele denne kunnskapen og øke produktiviteten i byggebransjen på en bærekraftig måte."

- **Dr. Michael Max Buehler**, professor i byggeledelse ved HTWG Konstanz, medeier i GemeinWerk Ventures og uavhengig direktør i DevvStream.

"DataDrivenConstruction-boken er et av de første skrittene utover grensene for den vanlige verdenen til byggherrer, med sine komplekse design- og styringssystemer, når det ser ut til at kompleksiteten og metningen av data ikke engang gir en sjanse for radikal forenkling og økt åpenhet i arbeidet med konstruksjonsdata. I sin bok viser Artem på et enkelt språk hvilke muligheter moderne teknologier for å jobbe med data som er åpne for oss, og gir bokstavelig talt konkrete trinn som du umiddelbart kan bruke i arbeidet ditt. Jeg oppfordrer alle som ønsker å forstå hvor automatiseringssystemene vil ta veien i byggebransjen, til å studere denne boken nøyde for å innse at datarevolusjonen i byggebransjen allerede banker på døren vår. Nå er det bare interessant for nerdene, men om noen få år vil slike tilnæringer og programvare være allestedsnærværende, i likhet med BIM!"

- **Ihor Rogachew**, leder for IMT Competence Centre, BIM & Digital Transformation hos RGD, og grunnlegger av InfraBIM.Pro.

"Jeg anbefaler DataDrivenConstruction-boken på det varmeste, som tar for seg, som tittelen sier, en datadrevet informasjonsstyringsmetode for AECO. Jeg bruker den for tiden til å sette i gang en rekke diskusjoner med ulike grupper. Jeg synes den er en svært tilgjengelig referanse. I tillegg til en grundig oversikt over den historiske konteksten for verktøy i AECO, data og introduksjon av flere nøkkleteknologier, inneholder boken en rekke svært nyttige diagrammer som skisserer omfanget av datakilder og sluttbrukerartefakter med eksempler på arbeidsflyter. Det slår meg at det er denne typen diagrammer vi trenger flere av når vi utvikler og overvåker informasjonsstrategier og bidrar til BEP-er - som definerer den overordnede datamodellen for virksomheten, og som kan brukes som ramme for en PIM og AIM."

- **Paul Ransley**, sjefskonsulent hos Acmena og systemintegreringsingenør hos Transport for London.

"Hvis "data er den nye oljen", må vi lære oss å definere den, finne den, utvinne den og foredle den for å gjøre den verdifull. Jeg har funnet boken DataDrivenConstruction svært informativ og innsiktsfull. Boken gir en nyttig historisk bakgrunn og forklarer hvordan man jobber med data på et enkelt språk. For de som er interessert i digital transformasjon, gir den en god forståelse av data - hvordan de fungerer, hvordan de er strukturert og hvordan de kan brukes."

- **Ralph Montague**, direktør i ArcDox, direktør for BIM Coordinators Summit og leder av BIM National Mirror Committee ved National Standards Authority of Ireland.

"Som det understres i boken, er informasjon en avgjørende ressurs for bygg- og anleggssektoren, og det å ha den i tilgjengelige formater gjør det lettere å ta riktige beslutninger og fremskynde prosjektene tidslinjer. Boken tilbyr en nøytral og effektiv tilnærming til å få tilgang til og dra nytte av denne kilden i beslutningsprosesser. Metoden som presenteres i boken, bygger på en moderne tilnærming som kombinerer programmering basert på kunstig intelligens med tilgjengelige verktøy med åpen kildekode. Ved å utnytte kraften i AI og bruke programvare med åpen kildekode tar metodikken sikte på å forbedre automatiseringen, optimalisere prosesser og fremme tilgjengelighet og samarbeid innen fagfeltet. Språket i boken er klart og enkelt å følge."

- **Dr. Salih Ofluoğlu**, dekan ved fakultetet for kunst og arkitektur ved Antalya Bilim University, og arrangør av Eurasian BIM Forum.

"Alt jeg kan si er: WOW! Måten du inkorporerte historie, LLM, grafikken og den generelle lettheten til å forstå poengene dine er virkelig bemerkelsesverdig. Flyten i boken er fantastisk. Det er så mange strålende aspekter ved denne boken; det er virkelig en spillveksler. Det er en flott kilde til informasjon, og jeg berømmer deg for innsatsen og lidenskapen du har lagt ned i den. Gratulerer med å ha skapt et så bemerkelsesverdig verk. Jeg kunne fortsette, men det er nok å si at jeg er utrolig imponert!"

- **Natasha Prinsloo**, Digital Practice Lead hos energylab_

"For alle i byggebransjen, fra nybegynnere til erfarte proffere, er denne boken en "game-changer"! Det er ikke en typisk støvete bok - den er fullpakket med innsikt, strategier og et snev av humor som holder deg engasjert. Boken tar for seg utviklingen av databruk i byggebransjen, fra eldgamle dataregisteringsmetoder til banebrytende digital teknologi. Det er som å ta en tidsmaskin gjennom utviklingen av byggdata. Enten du er arkitekt, ingeniør, prosjektleder eller dataanalytiker, vil denne omfattende guiden revolusjonere måten du tilnærmer deg prosjekter på. Gjør deg klar til å optimalisere prosesser, forbedre beslutningsprosesser og styre prosjekter som aldri før!"

- **Pierpaolo Vergati**, foreleser ved Sapienza-universitetet i Roma og senior prosjektleder for byggeprosjekter hos Fintecna.

"Jeg leste boken i ett pust, på mindre enn 6 timer. Bokens produksjonskvalitet er utmerket, tett glanset papir, fargevalg og en behagelig skriftype. Det store antallet praktiske eksempler på hvordan du kan jobbe med LLM spesifikt for byggebransjen vil spare deg for måneder, om ikke år, med selvstudier. Arbeidseksemplene er svært varierte, alt fra enkle til komplekse, uten at du trenger å kjøpe kompleks og kostbar programvare. Boken vil gi eiere av alle virksomheter i byggebransjen mulighet til å se på forretningsstrategi, digitalisering og utviklingsmuligheter med nye øyne. Og for mindre selskaper å øke effektiviteten med rimelige og gratis verktøy."

- **Mikhail Kosarev**, foreleser og konsulent innen digital transformasjon i byggebransjen ved TIM-ASG.

"Boken "DATA DRIVEN CONSTRUCTION" er en "game-changer" for alle som er nysgjerrige på hvor byggebransjen er på vei i dataalderen. Artem skraper ikke bare i overflaten; han går dypt inn i den aktuelle utviklingen, utfordringene og de lovende mulighetene i byggebransjen. Det som skiller denne boken fra andre, er dens tilgjengelighet - Artem forklarer komplekse ideer ved hjelp av relaterbare analogier som gjør innholdet lett å forstå. Jeg synes boken er utrolig informativ og samtidig engasjerende. Oppsummert har Artem laget en verdifull ressurs som ikke bare informerer, men også inspirerer. Enten du er en erfaren fagperson eller en nykommer i byggebransjen, vil denne boken utvide perspektivet ditt og gi deg en dypere forståelse av hvor bransjen er på vei. Anbefales på det sterkeste!"

- **Moayad Saleh**, arkitekt og BIM Implementation Manager hos TMM GROUP Gesamtplanungs GmbH.

"Jeg vil si at Data-Driven Construction fortjener å bli undervist i som lærebok på universitetene, og er en bok som vil gi verdifulle bidrag til utviklingen av BIM-feltet. Data-Driven Construction inneholder en teknisk ordliste som forklarer begrepene på en svært god måte. Temaer som er svært vanskelige å forklare, blir gjort enkle og forståelige med et veldig vakkert visuelt språk. Jeg mener at det som skal forklares i bildene, bør formidles til leseren, selv om det er kortfattet. Forståeligheten av noen bilder, med andre ord, krever separat informasjon for å lese det visuelle. Jeg vil også si at jeg gjerne introduserer Artem Boikos verdifulle arbeid i mine forelesninger og seminarer på universitetene."

- **Dr. Ediz Yazicioglu**, eier av ArchCube og foreleser i prosjektleidelse ved Institutt for arkitektur ved Istanbuls tekniske universitet og Medipol-universitetet.

"Data Driven Construction formidler på en levende måte det grunnleggende i informasjonsbasert arbeid med bygningsdata. En bok som tar for seg informasjonsflyt og grunnleggende økonomiske konsepter og dermed skiller seg ut fra andre BIM - bøker, fordi den ikke bare representerer perspektivet til en programvareprodusent, men også forsøker å formidle grunnleggende konsepter. En bok som er verdt å lese og se."

- **Jakob Hirn**, administrerende direktør og medgrunnlegger av Build Informed GmbH, og initiativtaker til innovasjonsforumet "On Top With BIM".

"Data er den nye oljen", som det heter, så de som leter etter den bør ha de rette verktøyene og den rette tankegangen for å kunne utvinne verdi fra denne ressursen i det 21. århundre. Byggebransjen har altfor lenge vært på et skråplan med prosesser basert på "3D-informasjon", der prosjektleveransen er basert på informasjon som noen andre har laget (f.eks. at de allerede har tegnet kake- eller søylediagrammet), mens de underliggende "dataene" (f.eks. det rå regnearket) kan levere mye mer, særlig fordi sammensmelting av flere data og kunstig intelligens gir ubegrensete muligheter. Hvis du leverer (eller underviser/forsker i) konstruksjon, er denne boken din beste - og hittil eneste - ressurs for å navigere i den datadrevne verdenen vi nå befinner oss i."

- **Dr. Zulfikar Adamu**, førsteamannuensis i strategisk IT i byggebransjen ved LSBU i Storbritannia.

"Data-Driven Construction" av Artem Boiko er et imponerende verk som gir et solid grunnlag for byggebransjen i en tid med stadig voksende teknologi og informasjonsmuligheter. Boiko klarer å presentere komplekse emner på en forståelig måte, samtidig som han introduserer visjonære ideer. Boken er et gjennomtenkt kompendium som ikke bare belyser dagens utvikling, men som også gir utsikter til fremtidige innovasjoner. Den anbefales på det varmeste for alle som ønsker å sette seg inn i datadrevet byggeplanlegging og -gjennomføring."

- **Markus Eiberger**, foreleser ved Stuttgart University of Applied Sciences, senior prosjektleder og nestleder i Konstruktionsgruppe Bauen, styremedlem i BIM Cluster Baden-Württemberg Association.



Andre utgave, mars 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-11-9



Artem Boiko Opphavsrett

boikoartem@gmail.com
info@datadrivenconstruction.io

Ingen del av denne boken kan reproduceres eller overføres i noen form eller på noen måte, elektronisk eller mekanisk, inkludert fotokopiering, opptak eller noe informasjonslagrings- og gjenfinningssystem, uten skriftlig tillatelse fra forfatteren - unntatt for ikke-kommersiell distribusjon i uendret form. Boken distribueres gratis og kan fritt deles med andre brukere for personlige, pedagogiske eller forskningsmessige formål, forutsatt at forfatterskap og referanser til originalen beholdes. Forfatteren beholder alle eiendomsrettigheter til teksten og gir ingen uttrykkelige eller underforståtte garantier. Selskaper, produkter og navn som nevnes i boken, kan være fiktive eller brukt som eksempler. Forfatteren er ikke ansvarlig for eventuelle konsekvenser av bruken av informasjonen som gis. Informasjonen i boken er gitt "som den er" uten garanti for fullstendighet eller relevans. Forfatteren er ikke ansvarlig for tilfeldige skader eller følgeskader som oppstår som følge av bruk av informasjonen, koden eller programmene i denne boken. Kodeeksemplene som presenteres i denne boken, er kun ment for undervisningsformål. Leserne bruker dem på egen risiko. Forfatteren anbefaler at alle programvarelösninger testes før de brukes i et produksjonsmiljø. Alle varemerker og produktnavn som nevnes i teksten, er varemerker, registrerte varemerker eller servicemerker for sine respektive selskaper og tilhører deres respektive eiere. Bruken av disse navnene i boken innebærer ikke noe forhold til eller støtte fra eierne. Omtale av produkter eller tjenester fra tredjeparter utgjør ikke en anbefaling eller innebærer en godkjenning. Firma- og produktnavn som brukes i eksemplene, kan være varemerker som tilhører eierne. Lenker til tredjeparts nettsteder er kun ment som en bekvemmelighet og innebærer ikke at forfatteren støtter informasjonen som gis på disse nettstedene. All statistikk, alle sitater og alle studier som er sitert, var gjeldende på det tidspunktet dette ble skrevet. Dataene kan endre seg over tid.

Denne boken er distribuert under en Creative Commons Navngivelse-Ikke-kommersiell-Ingen-Avleddede rettigheter 4.0 Internasjonal lisens (CC BY-NC-ND 4.0). Du kan kopiere og distribuere den for ikke-kommersielle formål, forutsatt at du beholder forfatterskapet og ikke gjør noen endringer.



CC BY-NC-ND 4.0

© 2024 Artjom Bojko. Første utgave.
© 2025 Artjom Bojko. Andre utgave, revidert og supplert.
Alle rettigheter forbeholdt.

FORORD TIL ANDRE UTGAVE

Denne boken er resultatet av en livlig dialog med fagmiljøet. Den er basert på en rekke faglige diskusjoner om datahåndtering i byggebransjen, som har funnet sted på ulike faglige plattformer og i sosiale medier. Disse diskusjonene har dannet grunnlag for artikler, publikasjoner og visuelt materiale som har skapt stor respons i fagmiljøet. Forfatterens innhold tiltrekker seg millioner av visninger hvert år på ulike plattformer og språk, og samler fagfolk innen digitalisering av bygg- og anleggsbransjen.

I løpet av ett år etter at den første utgaven kom ut, ble boken bestilt av eksperter fra mer enn 50 land, fra Brasil og Peru til Mauritius og Japan. Den andre utgaven av boken, som du nå holder i hendene, er revidert og utvidet på grunnlag av tilbakemeldinger fra eksperter, kritikk av den første utgaven og diskusjoner i fagmiljøer. Takket være tilbakemeldingene har andreutgaven blitt betydelig utvidet: nye kapitler om CAD (BIM) -teknologier og om å skape effektive ETL -prosesser er lagt til. Antallet praktiske eksempler og casestudier er også økt betraktelig. Av særlig verdi er tilbakemeldingene fra ledere i byggebransjen, konsulentsselskaper og store IT-selskaper, som har henvendt seg til forfatteren med spørsmål om digitalisering og interoperabilitet både før og etter at den første versjonen av boken ble publisert. Mange av dem har allerede tatt i bruk metodene som beskrives i boken, eller planlegger å gjøre det i nær fremtid.

Du holder i hendene en bok som er blitt til gjennom diskusjon og aktiv meningsutveksling. Fremskritt skapes i dialog, i brytningen av synspunkter og åpenhet for nye innfallsvinkler. Takk for at du er en del av denne dialogen. Din konstruktive kritikk er grunnlaget for fremtidige forbedringer. Hvis det er feil i teksten, eller hvis du ønsker å dele ideer og forslag, er alle tilbakemeldinger velkomne. Kontaktinformasjon finner du bakerst i boken.

HVORFOR ER BOKEN GRATIS?

Denne boken ble utviklet som en åpen utdanningsressurs med mål om å spre moderne tilnærminger til datahåndtering i byggebransjen. Den første versjonen av boken ble brukt som grunnlag for å samle inn kommentarer og forslag fra fagmiljøet, noe som gjorde det mulig å forbedre strukturen og innholdet i materialet. Alle kommentarer, forslag og ideer er nøye analysert og innarbeidet i denne reviderte versjonen. Målet med boken er å hjelpe fagfolk i byggebransjen til å forstå hvor viktig det er å arbeide med data: systematisk, bevisst og med tanke på den langsigtede verdien av informasjonen. Forfatteren har samlet eksempler, illustrasjoner og praktiske observasjoner fra mer enn 10 års arbeid med digitalisering av bygg og anlegg. Det meste av materialet er hentet fra reelle prosjekter, diskusjoner med ingeniører og byggmenn, deltakelse i internasjonale initiativer og opplæringsseminarer. Boken er et forsøk på å strukturere den samlede erfaringen og dele den på en tilgjengelig måte. Hvis du ønsker å støtte videre spredning av bokens ideer og få et praktisk format for å lese, jobbe med eksempler og visuelt materiale - kan du kjøpe en [trykt versjon](#).

BRUKSRETTIGHETER

Alt materiale, alle illustrasjoner og fragmenter i denne boken kan reproduceres, siteres eller brukes i et hvilket som helst format og på et hvilket som helst medium, forutsatt at kilden oppgis til Artem Boiko og at bokens tittel er "Data-Driven Construction". Takk for respekten for arbeid og kunnskapsformidling.

Det er med oppriktig takknemlighet jeg tilegner denne boken til familién min, som fra tidlig alder innpodet meg en dyp kjærlighet til bygg og anlegg, til gruvebyen der jeg bodde for å lære om motstandskraft, og til min landmålerhustru, hvis usvikelige støtte har vært min konstante inspirasjon.

HVEM DENNE BOKEN ER FOR

Denne boken er skrevet i et lettfattelig språk og henvender seg til et bredt spekter av lesere i byggebransjen - fra studenter og nybegynnere som ønsker å forstå det grunnleggende i moderne byggeprosesser, til fagfolk som trenger en oppdatert metodikk for datahåndtering i bygg- og anleggssbransjen. Enten du er arkitekt, ingeniør, formann, byggleder eller dataanalytiker, får du i denne omfattende guiden med mange unike illustrasjoner og grafikk verdifull innsikt i hvordan du kan bruke data i virksomheten til å optimalisere og automatisere prosesser, forbedre beslutningsprosesser og styre byggeprosjekter på ulike nivåer hjelpt av moderne verktøy.

Boken er en omfattende guide som kombinerer teoretisk grunnlag og praktiske anbefalinger for integrering av datahåndteringsteknikker i byggeprosesser. Boken fokuserer på strategisk bruk av informasjon for å optimalisere driften, automatisere prosesser, forbedre beslutningstaking og effektivt styre prosjekter ved hjelp av moderne digitale verktøy.

Denne boken tar for seg de teoretiske og praktiske aspektene ved å arbeide med informasjon i byggebransjen. Gjennom detaljerte eksempler utforskes metodikken for oppgaveparametrering, kravinnhenting, behandling av ustrukturerte data og data i flere formater, og hvordan disse kan omdannes til effektive løsninger for byggefirmaer.

Leseren går suksessivt fra utforming av krav og utvikling av grunnleggende datamodeller til mer komplekse prosesser for integrering av heterogene informasjonskilder, opprettelse av ETL -prosesser, konstruksjon av informasjonspipelines og maskinlæringsmodeller. Den sekvensielle tilnærmingen gjør det mulig å tydelig demonstrere mekanismene for organisering og automatisering av forretningsprosesser og beslutningsstøttesystemer i byggebransjen. Hver del av boken avsluttes med et praktisk kapittel med trinnvise instruksjoner som gjør det mulig å anvende den tilegnede kunnskapen umiddelbart i reelle prosjekter.

SAMMENDRAG AV DELER AV BOKEN

Denne boken er strukturert rundt konseptet datatransformasjon i verdikjeden: fra datainnsamling og kvalitetssikring til analytisk behandling og utvinning av verdifulle praktiske løsninger ved hjelp av moderne verktøy og metoder.

Del 1: Digital utvikling i bygg- og anleggsbransjen - spører den historiske transformasjonen av datahåndtering fra leirtavler til moderne digitale systemer, og analyserer fremveksten av modulære systemer og den økende betydningen av informasjonsdigitalisering i forbindelse med industrielle revolusjoner.

Del 2: Informasjonsutfordringer for byggebransjen - utforsker problemene med datafragmentering, "informasjonssiloer", HiPPO-tilnærmingens innvirkning beslutningstaking og begrensningene ved proprietære formater, og foreslår å vurdere overgangen til AI og LLM-økosystemer.

Del 3: Systematisering av data i bygg- og anleggsbransjen - danner en typologi for bygg- og anleggsdata, beskriver metoder for organisering og integrering med bedriftssystemer og diskuterer opprettelsen av kompetansesentre for standardisering av informasjonsprosesser.

Del 4: Kvalitetssikring av data - avslører metoder for å omdanne uensartet informasjon til strukturerte kvalitetsdata, inkludert datauttrekk fra ulike kilder, validering og modellering ved hjelp av LLM.

Del 5: Kostnads- og tidsberegninger - omhandler digitalisering av kostnads- og planleggingsberegnninger, automatisering av innhenting av volmer fra CAD-modeller (BIM), 4D-8D-modelleringsteknologier og ESG-beregning av byggeprosjekter.

Del 6: DAK og BIM - en kritisk analyse av utviklingen av designteknologier, problemer med interoperabilitet mellom systemer, trender i retning av åpne dataformater og utsiktene for bruk av kunstig intelligens i design.

Del 7: Dataanalyse og automatisering - tar for seg prinsippene for informasjonsvisualisering, nøkkelindikatorer, ETL-prosesser, verktøy for arbeidsflytorkestrering og bruk av språkmodeller for å automatisere rutineoppgaver.

Del 8: Datalagring og -administrasjon - utforsker datalagringsformater, datavarehus- og datasjøkonsepter, prinsipper for datahåndtering og nye tilnærninger, inkludert vektordatabaser og DataOps- og VectorOps-metodikkene.

Del 9: Big Data og maskinlæring - fokuserer på overgangen til objektiv analyse basert på historiske data, tingenes internett på byggeplasser og bruk av maskinlæringsalgoritmer for å forutsi prosjektkostnader og tidslinjer.

Del 10: Byggebransjen i de digitale dataenes tidsalder - tar for seg byggebransjens fremtid og analyserer overgangen fra kausalanalyse til arbeid med korrelasjoner, begrepet "Uberisering" av byggebransjen og strategier for digital transformasjon.

What is meant by **data-driven construction** ?



INNLEDNING

Hvor lenge kan bedriften din forbli konkurransedyktig i en verden der teknologien utvikler seg raskt og alle aspekter av virksomheten, fra timing og kostnadsberegning til risikoanalyse, blir automatisert av maskinlæringsmodeller?

Byggebransjen, som har eksistert like lenge som menneskeheden selv, står på terskelen til revolusjonerende endringer som vil forandre måten vi tenker på når det gjelder tradisjonell bygging. I andre deler av økonomien er digitaliseringen ikke bare i ferd med å endre spillereglene, men også med å drive ut av markedet selskaper som ikke har klart å tilpasse seg det nye databehandlingsmiljøet og ikke er i stand til å forbedre beslutningshastigheten (fig. 1).



Fig. 1 Beslutningshastigheten i byggebransjen er i større grad enn i andre bransjer avhengig av den menneskelige faktoren.

Bankvirksomhet, detaljhandel, logistikk og landbruk er på rask vei mot full digitalisering, der unøyaktigheter og subjektive meninger ikke lenger har noen plass. Moderne algoritmer er i stand til å analysere enorme datamengder og gi kundene nøyaktige prognosenter - enten det gjelder sannsynligheten for tilbakebetaling av lån, optimale leveringsruter eller risikopronosenter.

Bygg- og anleggsbransjen er en av de siste bransjene som er i ferd med å gå over fra løsninger basert på høyt betalte spesialisters meninger til datadrevne løsninger. Denne overgangen drives ikke bare av nye teknologiske muligheter, men også av markedets og kundenes økte krav til åpenhet, nøyaktighet og hastighet.

Robotisering, prosessautomatisering, åpne data og prognosenter basert på dem - alt dette er ikke lenger bare muligheter, men uunngåelig. De fleste selskaper i byggebransjen, som nylig var ansvarlige overfor kunden for å beregne volum, kostnader, prosjekttid og kvalitetskontroll, risikerer nå å bli rene utførere av ordrer, uten å ta viktige beslutninger (fig. 2).

Med fremskritt innen datakraft, maskinlæringsalgoritmer og demokratisering av datatilgang er det nå mulig å kombinere data fra ulike kilder automatisk, noe som muliggjør dypere prosessanalyse, risikoprediksjon og kostnadsoptimalisering i diskusjonsfasen av et byggeprosjekt. Disse teknologiene har potensial til å forbedre effektiviteten og redusere kostnadene radikalt i hele sektoren.

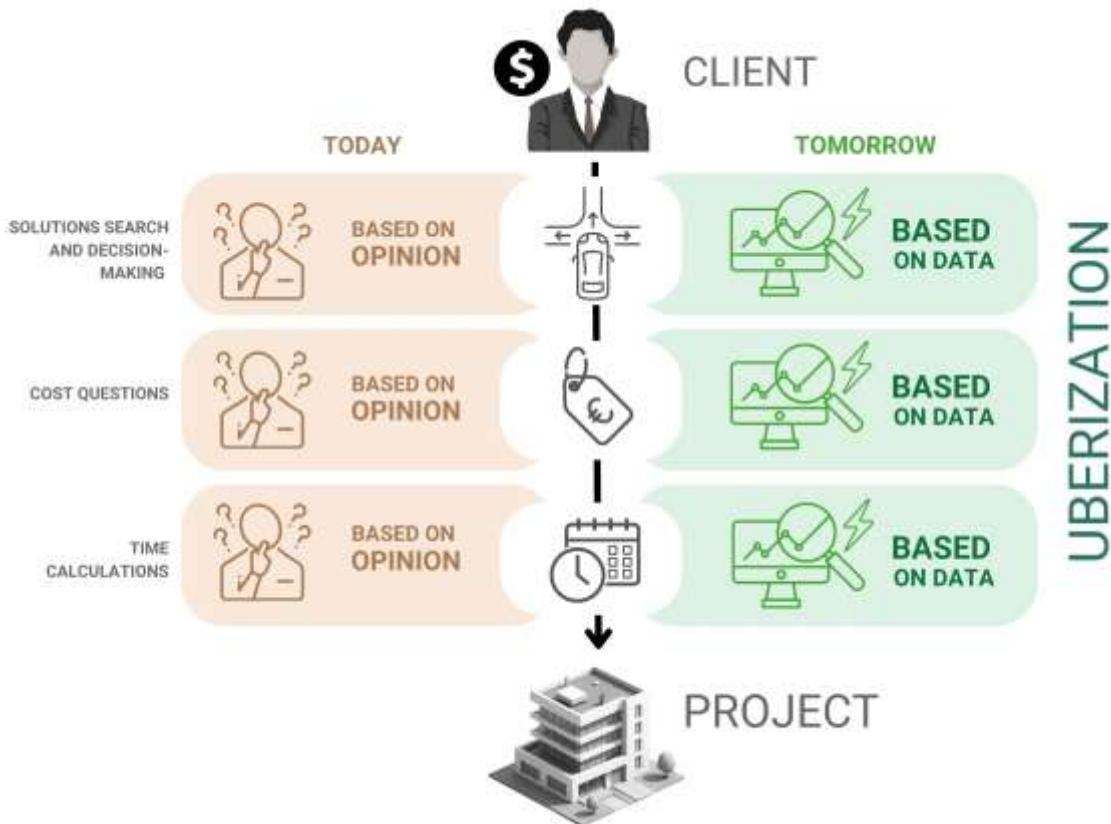


Fig. 2 Kunden er ikke interessert i for mange menneskelige faktorer på veien mot realisering av prosjektet.

Til tross for alle fordelene med nye verktøy og konsepter, ligger byggebransjen langt etter andre sektorer i økonomien når det gjelder å ta i bruk ny teknologi.

Ifølge rapporten IT Metrics Key Data 2017 ligger byggebransjen på siste plass blant 19 andre bransjer når det gjelder IT-investeringer [1].

Den raske veksten i datamengde og prosesskompleksitet er i ferd med å bli en hodepine for bedriftsledelsen, og hovedproblemet ved bruk av ny teknologi er at data, til tross for at de finnes i overflod, fortsatt er fragmenterte, ustrukturerte og ofte inkompatible mellom ulike systemer og programvareprodukter. Derfor er mange selskaper i bygg- og anleggsbransjen nå først og fremst opptatt av datakvalitetsproblemer, som bare kan løses ved å ta i bruk effektive, automatiserte styrings- og analysesystemer.

Ifølge en undersøkelse gjennomført av KPMG® blant byggeledere i 2023 [2] har informasjonsystemer for projektstyring (PMIS), avansert og grunnleggende dataanalyse og bygningsinformasjonsmodellering (BIM) det største potensialet for å forbedre prosjektenes avkastning på investeringen (figur 3).

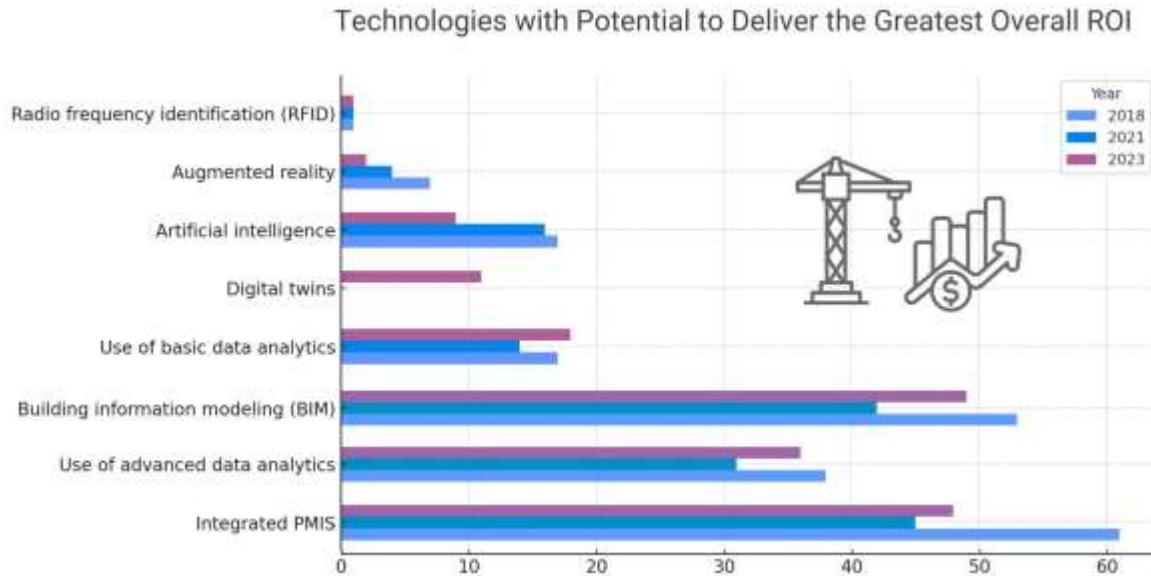


Fig. 3 Spørreundersøkelse blant ledere i byggefirmaer: Hvilke teknologier vil gi høyest avkastning på investeringen (ROI) i kapitalprosjekter? (basert på materiale [2]).

Løsningen på utfordringene knyttet til integrering av data i forretningsprosesser er å sikre høy kvalitet på informasjonen, bruke egnede dataformater og anvende effektive metoder for å skape, lagre, analysere og behandle data.

Erkjennelsen av verdien av data tvinger ulike bransjer til å bevege seg bort fra siloapplikasjoner og komplekse byråkratiske styringsstrukturer. I stedet flyttes fokuset over på å skape nye tilnærminger til informasjonsarkitektur, slik at selskapene forvandles til moderne datadrevne virksomheter. Før eller siden vil byggebransjen selv ta dette skrittet, og gå fra en gradvis digital utvikling til en ekte digital revolusjon som påvirker alle selskaper.

Overgangen til datadrevne forretningsprosesser kommer ikke til å bli enkel. Mange bedrifter vil møte utfordringer fordi lederne ikke alltid forstår hvordan de kan bruke kaotiske datasett til å forbedre effektiviteten og veksten i virksomheten.

Denne boken dykker ned i en verden av data, der informasjon er i ferd med å bli en strategisk nøkkelressurs som avgjør effektiviteten og bærekraften i forretningsprosesser. Med den raske veksten i informasjonsmengden står bedriftene overfor nye utfordringer. Digital transformasjon er ikke lenger bare et moteord - det er i ferd med å bli en nødvendighet.



Fig. 4 Data og prosesser er fundamentet for bygg og anlegg.

Å forstå transformasjon betyr å kunne forklare det komplekse med enkle ord. Derfor er boken skrevet i et lettfattelig språk og ledsaget av forfatterens egne illustrasjoner, som er laget spesielt for å forklare sentrale begreper på en tydelig måte. Disse diogrammene, grafiske fremstillingene og visualiseringene er utformet for å fjerne persepsjonsbarrierer og gjøre stoffet forståelig også for dem som tidligere anså slike emner som for komplekse. Alle illustrasjoner, diagrammer og grafikk i denne boken er laget av forfatteren og er spesielt utformet for å visualisere de sentrale begrepene som beskrives i teksten.

Et bilde sier mer enn tusen ord [3].

- Fred R. Barnard, engelsk illustratør, 1927.

For å knytte teori og praksis sammen bruker vi verktøy for kunstig intelligens (særlig språkmodeller) som gjør det mulig å utvikle løsninger uten å måtte ha inngående programmeringskunnskaper. Hvis du er orientert mot praktisk materiale og er mer interessert i praktisk arbeid med data, kan du hoppe over den første introduksjonsdelen og gå rett til den andre delen av boken, der beskrivelsen av konkrete eksempler og caser starter.

Men ikke ha for store forventninger til AI (kunstig intelligens), maskinlæring og LLM-verktøy (Large Language Models) generelt. Uten gode inngangsdata og en dyp forståelse av emnet kan selv de mest avanserte algoritmene ikke gi pålitelige og meningsfulle resultater.

Microsofts administrerende direktør Satya Nadella advarer mot risikoen for en boble innen kunstig intelligens tidlig i 2025 [4], og sammenligner den nåværende hypen med dot-com-boblen. Han understreker at påstander om at man har nådd milepæler innen AGI (Artificial General Intelligence) uten en skikkelig begrunnelse, er "meningsløs manipulering av måleparametere". Nadella mener at den virkelige suksessen til AI bør måles ut fra bidraget til global BNP-vekst, i stedet for å legge for stor vekt på moteord.

Bak alle moteordene om ny teknologi og nye konsepter ligger det et komplekst og møysommelig arbeid

med å sikre datakvalitet, parametrisering av forretningsprosesser og tilpasning av verktøy til reelle oppgaver.

En datadrevet tilnærming er ikke et produkt du bare kan laste ned eller kjøpe. Det er en strategi som må bygges opp. Den starter med et nytt blikk på eksisterende prosesser og problemer, og krever deretter en disciplinert bevegelse i den valgte retningen.

Ledende programvareutviklere og applikasjonsleverandører vil ikke være motoren for endring i byggebransjen, for mange av dem er en datadrevet tilnærming en trussel mot deres etablerte forretningsmodell.

Andre bransjer [i motsetning til byggebransjen], som bilindustrien, har allerede gjennomgått radikale og disruptive endringer, og den digitale transformasjonen er godt i gang. Bygg- og anleggsbransjen må handle raskt og beslutsomt: De som er raske, vil høste store gevinst, mens de som nøler, vil risikere mye. Tenk bare på omveltingene som digital fotografering har forårsaket i denne bransjen [5].

- World Economic Forums rapport Shaping the Future of Construction, 2016

De selskapene som i tide ser mulighetene og fordelene ved den nye tilnærmingen, vil få et varig konkurransefortrinn og vil kunne utvikle seg og vokse uten å være avhengige av løsninger fra store leverandører.

Dette er din sjanse til ikke bare å ri av den kommende stormen med digitalisering av informasjon, men til å ta kontroll over den. I denne boken finner du ikke bare en analyse av den nåværende tilstanden i bransjen, men også konkrete anbefalinger for hvordan du kan tenke nytt og omstrukturere prosessene og virksomheten din for å bli ledende i den nye epoken i byggebransjen og forbedre din yrkesfaring.

Den digitale fremtiden i byggebransjen handler ikke bare om å ta i bruk ny teknologi og nye programmer, men også om å tenke fundamentalt nytt når det gjelder datahåndtering og forretningsmodeller.

Er bedriften din klar for denne strategiske endringen?

INNHOLDSFORTEGNELSE

INNLEDNING	1
INNHOLDSFORTEGNELSE	I
II DEL FRA LEIRTAVLER TIL DEN DIGITALE REVOLUSJONEN: HVORDAN INFORMASJON HAR UTVIKLET SEG I BYGGEBRANSJEN	2
KAPITTEL 2.1. UTVIKLINGEN AV DATABRUK I BYGGEBRANSJEN	3
Dataåraens fødsel i byggebransjen.....	3
Fra leire og papyrus til digital teknologi.....	4
Prosess som verktøy for datadrevne opplevelser.....	5
Digitalisering av informasjon om byggeprosessen	7
KAPITTEL 2.2. TEKNOLOGIER OG STYRINGSSYSTEMER I MODERNE BYGG OG ANLEGG.....	11
Den digitale revolusjonen og fremveksten av modulære MRP/ERP -systemer.....	11
Datahåndteringssystemer: fra datautvinning til forretningsutfordringer	13
Mycelium i bedriften: hvordan data kobles til forretningsprosesser	16
KAPITTEL 2.3. DEN DIGITALE REVOLUSJONEN OG EKSPLOSIONEN AV DATA.....	20
Begynnelsen på databoomen som en evolusjonær bølge	20
Mengden data som genereres i en moderne bedrift	22
Kostnadene ved datalagring: det økonomiske aspektet.....	23
Grenser for dataakkumulering: fra masse til mening	25
Neste steg: fra datateori til praktisk endring	26
III DEL HVORDAN BYGGEBRANSJEN DRUKNER I DATAKAOS.....	28
KAPITTEL 3.1. DATAFRAGMENTERING OG SILOER	29
Jo flere verktøy, desto mer effektiv blir virksomheten?	29
Datasiloer og deres innvirkning på selskapets resultater	31
Duplisering og manglende datakvalitet som en konsekvens av uenighet	34
HiPPO eller faren for meninger i beslutningsprosesser	36
Kontinuerlig økning i kompleksiteten og dynamikken i forretningsprosessene	39
Den fjerde industrielle revolusjonen (Industri 4.0) og den femte industrielle revolusjonen (Industri 5.0) i bygg- og anleggsbransjen.....	42
KAPITTEL 3.2. GJØRE KAOS TIL ORDEN OG REDUSERE KOMPLEKSITETEN	45
Overflødig kode og lukkede systemer hindrer produktivitetsforbedringer	45
Fra siloer til ett enkelt datavarehus	47

Integrerte lagringssystemer muliggjør overgangen til AI agenter	48
Fra datainnsamling til beslutningstaking: veien mot automatisering.....	51
Neste steg: Gjør kaoset om til et håndterbart system.....	53
IV DEL DATARAMMEVERK I FORRETNINGSPROSesser I BYGG- OG ANLEGGsBRANSJEN55	
KAPITTEL 4.1. DATATyPER I BYGG- OG ANLEGGsBRANSJEN	56
De viktigste datatypene i byggebransjen.....	56
Strukturerte data.....	60
Relasjonsdatabaser RDBMS og spørrespråket SQL.....	61
SQL - forespørsler i databaser og nye trender	64
Ustrukturerte data	66
Tekstdata: mellom ustrukturert kaos og strukturert kaos y	67
Semistrukturerte og løst strukturerte data.....	68
Geometriske data og anvendelse.....	69
CAD data: fra design til datalagring.....	72
Fremveksten av BIM-konseptet (BOM) og bruk av CAD i prosesser.....	75
KAPITTEL 4.2. HARMONISERING OG STRUKTURERING AV DATA	81
Fyller systemer med data i byggebransjen	81
Datatransformasjon: det kritiske grunnlaget for moderne forretningsanalyse	83
Datamodeller: relasjoner i dataene og relasjoner mellom elementer	87
Proprietære formater og deres innvirkning på digitale prosesser.....	91
Åpne formater endrer tilnærmingen til digitalisering.....	95
Paradigmeskifte: Åpen kildekode som slutten på epoken med programvareleverandørers dominans	96
Strukturerte åpne data: grunnlaget for digital transformasjon	98
KAPITTEL 4.3. LLM OG DERES ROLLE I DATABEHANDLING OG FORRETNINGSPROSesser	102
LLM chatterom: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok for å automatisere dataprosesser.....	102
Store LLM-språkmodeller: hvordan det fungerer	103
Bruk av lokale LLM-er for sensitive bedriftsdata	106
Full kontroll over AI i selskapet og hvordan du distribuerer din egen LLM	108
RAG: Intelligent LLM -assisterter med tilgang til bedriftsdata	110
KAPITTEL 4.4. IDE MED LLM-STØTTE OG FREMTIDIGE PROGRAMMERINGSENDRINGER	112
Valg av IDE: fra LLM eksperimenter til forretningsløsninger	112
IDE med LLM-støtte og fremtidige programmeringsendringer	114

Python Pandas: et uunnværlig verktøy for arbeid med data.....	115
DataFrame: universelt tabellformat for data.....	119
Neste steg: å bygge et bærekraftig datarammeverk	122
V DEL DATAKVALITET: ORGANISERING, STRUKTURERING, MODELLERING	124
KAPITTEL 5.1. KONVERTERING AV DATA TIL EN STRUKTURERT FORM	125
Lære hvordan du gjør dokumenter, PDF-filer, bilder og tekster om til strukturerte formater	125
Eksempel på konvertering av et PDF -dokument til en tabell	126
Konverter JPEG- og PNG-bilder til strukturert form	130
Konvertering av tekstdata til en strukturert form	132
Konvertering av CAD-data (BIM) til en strukturert form	135
CAD - leverandørene av løsninger går i retning av strukturerte data.....	139
KAPITTEL 5.2. KLASSIFISERING OG INTEGRERING: ET FELLES SPRÅK FOR BYGGDATA	143
Hvor raskt beslutningene tas, avhenger av kvaliteten på dataene	143
Datastandardisering og -integrasjon.....	144
Digital interoperabilitet starter med krav.....	146
Et felles konstruksjonsspråk: Klassifiseringsverktøyenes rolle i den digitale transformasjonen.....	149
Masterformat, OmniClass, Uniclass og CoClass: utviklingen av klassifikasjonssystemer	151
KAPITTEL 5.3. DATAMODELLERING OG KOMPETANSEENTER	156
Datamodellering: konseptuell, logisk og fysisk modell.....	156
Praktisk datamodellering i forbindelse med bygg- og anleggsvirksomhet	160
Opprette en database ved hjelp av LLM	162
Senter for fremragende forskning (CoE) for datamodellering.....	164
KAPITTEL 5.4. SYSTEMATISERING AV KRAV OG VALIDERING AV INFORMASJON	167
Innhenting og analyse av krav: omforming av kommunikasjon til strukturerte data	167
Prosessflytskjemaer og effektiviteten til konseptuelle rammeverk	171
Strukturerte krav og regulære RegEx-uttrykk.....	173
Datainnsamling for verifiseringsprosessen	178
Verifisering av data og resultater av verifiseringen	180
Visualisering av verifiseringsresultater på	185
Sammenligning av datakvalitetssjekker med menneskers livsbehov	187
Neste trinn: gjøre data om til nøyaktige beregninger og planer	189
VI DEL KOSTNADS- OG TIDSBEREGNINGER: INNLEMME DATA I BYGGEPROSESSEN	191
KAPITTEL 6.1. KOSTNADSBEREGNINGER OG ESTIMATER FOR BYGGEPROJEKTER	192
Grunnleggende om bygging: estimering av mengde, kostnad og tid	192

Metoder for beregning av estimerte prosjektkostnader	193
Ressursbasert metode kostnadsberegning og estimering i bygg- og anleggsbransjen	194
Database over byggeressurser: katalog over byggematerialer og byggverk	194
Utarbeidelse av kalkyler og kostnadsberegning av arbeider basert på ressursgrunnlaget	196
Sluttkalkulasjon av prosjekter: fra estimator til budsjetter	200
KAPITTEL 6.2. MENGDEBEREGNING OG AUTOMATISK OPPRETTING AV ESTIMATER OG TIDSPLANER	
205	
Fra 3D til 4D og 5D: bruk av volumetriske og kvantitative parametere.....	205
5D-attributter og innhenting av attributtvolum fra CAD	205
QTO Quantity Take-Off: gruppering av prosjektdata etter attributter.....	209
QTO-automatisering ved hjelp av LLM og strukturerte data	214
QTO beregning av hele prosjektet ved hjelp av grupperegler fra et Excel-regneark	218
KAPITTEL 6.3. 4D, 6D -8D OG BEREGNING AV KARBONDIOKSIDUTSLIPP CO₂	224
4D-modell: Integrering av tid i konstruksjonsestimater	224
Byggeplan og automatisering av denne basert på kostnadsdata	225
Utvidede attributtlag 6D -8D: fra energieffektivitet til sikkerhetssikring	227
Estimering av CO ₂ og beregning av karbondioksidutslipp fra byggeprosjekter	230
KAPITTEL 6.4. CONSTRUCTION ERP OG PMIS SYSTEMER	234
Construction ERP -systemer på eksemplet med kalkyler og estimator	234
PMIS: Mellomledd mellom ERP og byggeplassen	239
Spekulasjon, profit, isolasjon og mangel på åpenhet i ERP og PMIS	240
Slutten på epoken med lukket ERP /PMIS: byggebransjen trenger nye tilnærmingar	242
Neste steg: effektiv bruk av prosjektdata	244
VII DEL CAD OG BIM: MARKEDSFØRING, VIRKELIGHET OG FREMTIDEN FOR DESIGNDATA I BYGG- OG ANLEGGSSBRANSJEN	247
KAPITTEL 7.1. FREMVEKSTEN AV BIM- KONSEPTER I BYGGEBRANSJEN	248
Historien bak fremveksten av BIM og åpen BIM som markedsføringskonsepter for CAD- leverandører.....	248
Virkeligheten i BIM: i stedet for integrerte databaser - lukkede modulære systemer.....	251
Fremveksten av det åpne formatet IFC i byggebransjen	253
IFC formatproblem avhengig av den geometriske kjernen.....	255
Utseende i konstruksjonen av emnet semantikk og ontologi.....	258
Hvorfor semantiske teknologier ikke lever opp til forventningene i byggebransjen.....	260
KAPITTEL 7.2. LUKKEDE PROSJEKTFORMATER OG PROBLEMER MED INTEROPERABILITET	263
Lukkede data og fallende produktivitet: CAD-bransjens blindvei (BIM).....	263

Myten om interoperabilitet mellom CAD-systemer	265
Gå til USD og detaljerte data m	269
KAPITTEL 7.3. GEOMETRI I BYGG OG ANLEGG: FRA LINJER TIL KUBIKKMETER	273
Når linjer blir til penger eller hvorfor byggherrer trenger geometri	273
Fra linjer til volumer: Hvordan areal og volum blir til data.....	273
Overgang til MESH, USD og polygoner: bruk av tessellering for geometri.....	275
LOD, LOI, LOMD - unik klassifisering av detaljer i CAD (BIM)	277
Nye CAD-standarder (BIM) - AIA, BEP, IDS, LOD, COBie	280
KAPITTEL 7.4. DESIGNPARAMETRISERING OG BRUK AV LLM FOR CAD-DRIFT	284
Illusjonen om at CAD-data er unike (BIM): veien til analyse og åpne formater	284
Design gjennom parametere: fremtiden for CAD og BIM.....	287
Fremveksten av LLM i design CAD databehandlingsprosesser	290
Automatisert analyse av DWG -filer med LLM og Pandas	293
Neste skritt: fra lukkede formater til åpne data	299
VIII DEL DATAREVET BESLUTNINGSTAKING, ANALYSE, AUTOMATISERING OG MASKINLÆRING .. 301	
KAPITTEL 8.1. DATAANALYSE OG DATADREVET BESLUTNINGSTAKING	302
Data som en ressurs i beslutningsprosessen	302
Visualisering av data: nøkkelen til forståelse og beslutningstaking	305
KPI-er og ROI	307
Dashboard og instrumentpaneler: Visualisering av måltall for effektiv ledelse.....	309
Dataanalyse og kunsten å stille spørsmål	311
KAPITTEL 8.2. DATAFLYT UTEN MANUELL INNSATS: HVORFOR ETL ER NØDVENDIG.....	314
ETL-automatisering: reduserte kostnader og raskere datahåndtering	314
ETL Utdrag: datainnsamling	317
ETL Transform: anvendelse av validerings- og transformasjonsregler	320
ETL Load: Visualiser resultater i diagrammer og grafer	323
ETL Load: Automatisk opprettelse av PDF-dokumenter	328
ETL Load: automatisk generering av dokumenter fra FPDF.....	329
ETL Load: Rapportering og innlasting til andre systemer	333
ETL med LLM: Visualisering av -data fra PDF -dokumenter	334
KAPITTEL 8.3. AUTOMATISK ETL TRANSPORTBÅND (RØRLEDNING).....	339
Pipeline: Automatisk ETL transportbånd data	339
Pipeline -ETL datavalideringsprosess med LLM	343
Rørledning -ETL: verifikasiing av data og informasjon om prosjektelementer i CAD (BIM)	345

KAPITTEL 8.4. ORKESTRERING AV ETL OG ARBEIDSEFLYTER: PRAKTISKE LØSNINGER	351
DAG og Apache Airflow: automatisering og orkestrering av arbeidsflyt	351
Apache Airflow: praktisk anvendelse ETL-automatisering	352
Apache NiFi for ruting og datakonvertering	356
n8n Low-Code, No-Code prosessorkestrering	357
Neste steg: fra manuelle operasjoner til analysebaserte løsninger	360
IX DEL LAGRING OG HÅNDTERING AV DATA I BYGG- OG ANLEGGSSBRANSJEN	362
KAPITTEL 9.1. DATAINFRASTRUKTUR: FRA LAGRINGSFORMATER TIL DIGITALE REPOSITORIER.....	363
Dataatomer: grunnlaget for effektiv informasjonsforvaltning	363
Lagring av informasjon: filer eller data	364
Lagring av stordata: Analyse av populære formater og deres effektivitet	366
Optimaliser lagring med Apache Parquet	368
DWH: Data Warehouse datavarehus	370
Data Lake - utviklingen fra ETL til ELT: fra tradisjonellrensing til fleksibel behandling	371
Data Lakehouse-arkitektur: synergimellom datalagre og datasjøer	373
CDE, PMIS, ERP eller DWH og Data Lake	375
KAPITTEL 9.2. DATAVAREHUSADMINISTRASJON OG FOREBYGGING AV KAOS.....	379
Vektordatabaser og den avgrensende boksen	379
Data Governance), Dataminimalisme) og Data Swamp)	382
DataOps og VectorOps: nye datastandarder	385
Neste steg: fra kaotisk lagring til strukturert lagring	387
X DEL STORDATA, MASKINLÆRING OG PREDIKSJONER	389
KAPITTEL 10.1. STORDATA OG ANALYSE AV DISSE	390
Stordata i byggebransjen: fra intuisjon til forutsigbarhet	390
Spørsmål om gjennomførbarheten av stordata: korrelasjon, statistikk og datautvalg	391
Stordata: analyse av data fra San Franciscos datasett med en million byggetillatelser	394
Eksempel på stordata basert på CAD-data (BIM)	399
IoT Tingenes internett og smartkontrakter	403
KAPITTEL 10.2. MASKINLÆRING OG PREDIKSJONER	407
Maskinlæring og kunstig intelligens vil endre måten vi bygger på	407
Fra subjektiv vurdering til statistisk prognose	409
Titanic-datasettet: Hello World i analyseverdenen data og stordata	411
Maskinlæring i aksjon: fra Titanic-passasjerer til prosjektleddelse	416
Forutsigelser og prognoser basert på historiske data	420

Nøkkelsbegreper innen maskinlæring.....	422
KAPITTEL 10.3. FORUTSI KOSTNADER OG TIDSFRISTER VED HJELP AV MASKINLÆRING.....	425
Et eksempel på bruk av maskinlæring for å finne prosjektkostnader og tidslinjer	425
Forutsigelse av prosjektkostnader og -tid ved hjelp av lineær regresjon.....	427
Prognoser for prosjektkostnader og -tid ved hjelp av K-nærmeste naboalgoritmen (k-NN)	430
Neste steg: fra lagring til analyse og prognosering.....	434
XI DEL BYGGEVIRKSOMHETEN I DEN DIGITALE TIDSALDER. MULIGHETER OG UTFORDRINGER.....	437
KAPITTEL 11.1. OVERLEVELSESSTRATEGIER: Å BYGGE KONKURANSEFORTRINN.....	438
Korrelasjoner i stedet for beregninger: Fremtiden for bygganalyse	438
Datadrevet tilnærming i bygg- og anleggsbransjen: et nytt nivå av infrastruktur.....	441
Neste generasjons digitale kontor: Hvordan AI endrer arbeidsplassen	443
Åpne data og Uberisering er en trussel mot eksisterende bygg- og anleggsvirksomhet	445
Uløste problemer med uberisering som siste sjanse til å bruke tiden til transformasjon	448
KAPITTEL 11.2. EN PRAKTISK GUIDE TIL IMPLEMENTERING AV EN DATADREVET TILNÆRMING	453
Fra teori til praksis: Et veikart for digital transformasjon i byggebransjen.....	453
Å legge det digitale grunnlaget: 1-5 trinn til digital modenhet	455
Frigjør potensialet i data: 5-10 trinn til digital modenhet	459
Veikart for transformasjon: fra kaos til datadrevet selskap	465
Bygge i Industri 5.0: Hvordan tjene penger når du ikke kan gjemme deg lenger.....	468
KONKLUSJON	470
OM FORFATTEREN.....	473
VIDERE FORHOLD	474
OVERSETTELSESKOMMENTAR.....	474
ANDRE FERDIGHETER OG KONSEPTER	475
GLOSSAR.....	479
LISTE OVER REFERANSER OG NETTBASERT MATERIALE	485
INDEX FOR EMNE	501

MAKSIMAL BEKVEMMELIGHET MED DEN TRYKTE VERSJONEN

Du holder en gratis digital versjon av **Datadrevet konstruksjon**. For mer praktisk arbeid og rask tilgang til materialet anbefaler vi deg å ta hensyn til den **trykte utgaven**:



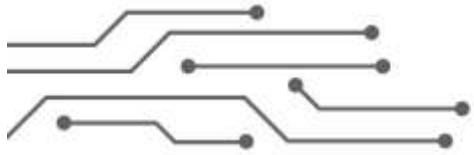
■ **Alltid tilgjengelig:** Boken i trykt format blir et pålitelig arbeidsverktøy, slik at du raskt kan finne og bruke de riktige visualiseringene og diagrammene i enhver arbeidssituasjon

■ **Illustrasjoner av høy kvalitet:** Alle bilder og all grafikk i den trykte utgaven presenteres i maksimal kvalitet

■ **Rask tilgang til informasjon:** enkel navigering, mulighet til å lage notater, bokmerker og arbeide med boken hvor som helst.

Ved å kjøpe den trykte versjonen av boken får du et praktisk verktøy for komfortabelt og effektivt arbeid med informasjon: muligheten til raskt å bruke visuelt materiale i daglige oppgaver, raskt finne de nødvendige ordningene og lage notater. I tillegg støtter kjøpet ditt formidling av åpen kunnskap.

Bestill en trykt versjon av boken på: datadrivenconstruction.io/books



I DEL

FRA LEIRTAVLER TIL DEN DIGITALE REVOLUSJONEN: HVORDAN INFORMASJON HAR UTVIKLET SEG I BYGGEBRANSJEN

Den første delen av boken tar for seg den historiske utviklingen av datahåndtering i byggebransjen, fra primitive registreringer på fysiske medier til moderne digitale økosystemer. Den analyserer transformasjonen av informasjonshåndteringsteknologier, fremveksten av ERP-systemer og virkningen av datafragmentering på effektiviteten i forretningsprosessene. Det legges særlig vekt på digitaliseringen av informasjon og den økende betydningen av objektiv analyse i stedet for subjektive ekspertervurderinger. Den eksponentielle veksten i informasjonsmengden som den moderne byggebransjen står overfor, og de tilhørende utfordringene for forretningssystemer, blir undersøkt i detalj. Vi ser nærmere på byggebransjens posisjonering i forbindelse med den fjerde og femte industrielle revolusjon, og på potensialet som ligger i å bruke kunstig intelligens og datasentriske tilnærmingar for å skape bærekraftige konkurransefortrinn.

KAPITTEL 1.1.

UTVIKLINGEN AV DATABRUK I BYGGEBRANSJEN

Dataæraens fødsel i byggebransjen

For rundt 10 000 år siden, i den neolittiske perioden, gjorde menneskeheten en revolusjonerende overgang i sin utvikling, da den forlot nomadetilværelsen til fordel for et bofast liv, noe som førte til at de første primitive bygningene av leire, tre og stein dukket opp [6]. Fra dette øyeblikket begynner byggebransjens historie.

Etter hvert som sivilisasjonene utviklet seg, ble arkitekturen stadig mer kompleks, noe som førte til de første rituelle templene og offentlige bygningene. De stadig mer komplekse arkitektoniske designene krevde at antikkens ingeniører og ledere måtte lage de første opptegnelsene og beregningene. De første nedtegnelsene på leirtavler og papyrus inneholder ofte en beskrivelse av logikken bak beregningen av hvor mye byggematerialer som trengs, hva de koster, og hvordan man beregner betalingen for det utførte arbeidet [7]. Slik begynte en æra med bruk av data i bygg- og anleggsbransjen - lenge før moderne digital teknologi ble tatt i bruk (figur 1.1-1).

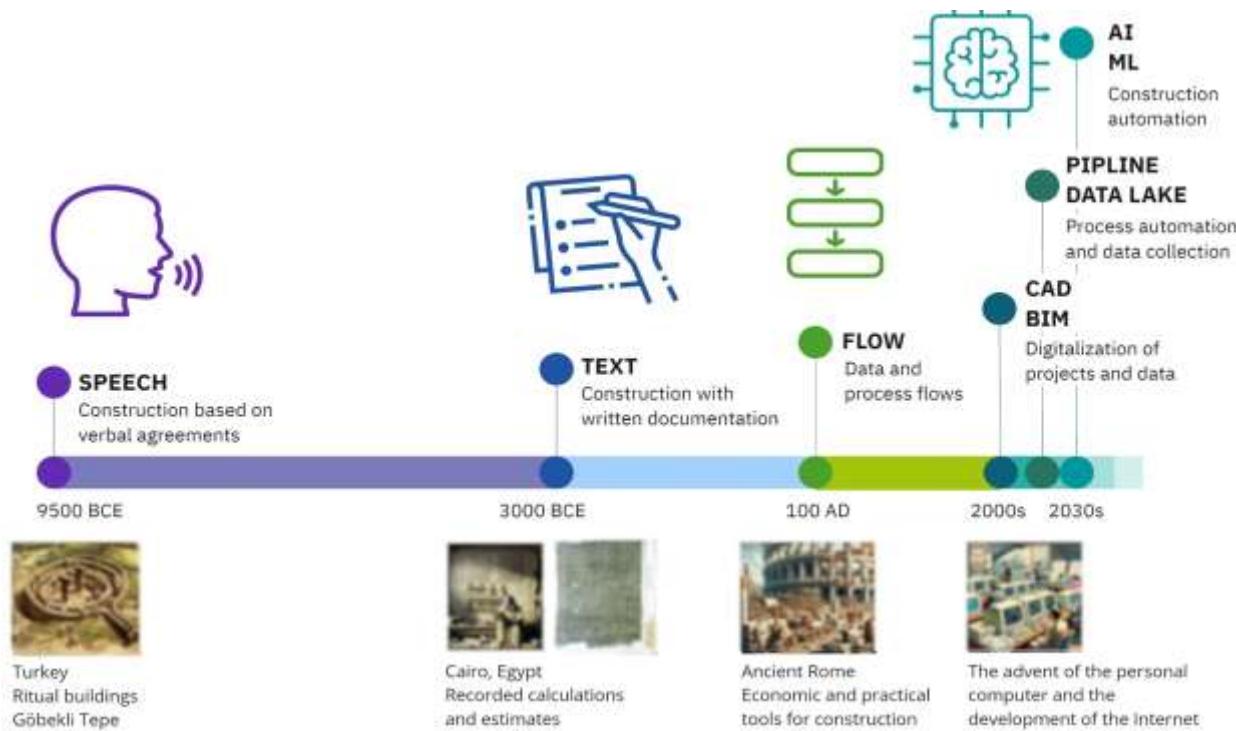


Fig. 1.1-1 Kronologi over informasjonsteknologiens utvikling i bygg- og anleggsbransjen: fra verbal informasjon til kunstig intelligens.

Fra leire og papyrus til digital teknologi

De første skriftlige bevisene innen bygg og anlegg stammer fra pyramidebyggingen, rundt 3000-4000 f.Kr.[7]. Siden den gang har skriftlig nedtegnelser lettet og ledsaget utviklingen i byggebransjen, og gjort det mulig å samle og systematisere kunnskap som i løpet av de neste 10 000 årene førte til betydelige innovasjoner innen byggemetoder og arkitektur.

Da de første fysiske mediene i bygg- og anleggsbransjen, som leirtavler, papyrus fra flere tusen år tilbake (figur 1.1-2) eller "A0"-papir på 1980-tallet, ble brukt til å registrere data, var det opprinnelig ikke meningen at denne informasjonen skulle brukes i nye prosjekter. Hovedformålet med slike nedtegnelser var å gi en detaljert oversikt over prosjektets nåværende status, inkludert beregninger av materialbehov og kostnader. På samme måte garanterer ikke tilgjengeligheten av digitale designdata og modeller i dagens verden alltid at de kan brukes i fremtidige prosjekter, og ofte fungerer de hovedsakelig som informasjon for aktuelle beregninger av materialbehov og byggekostnader.

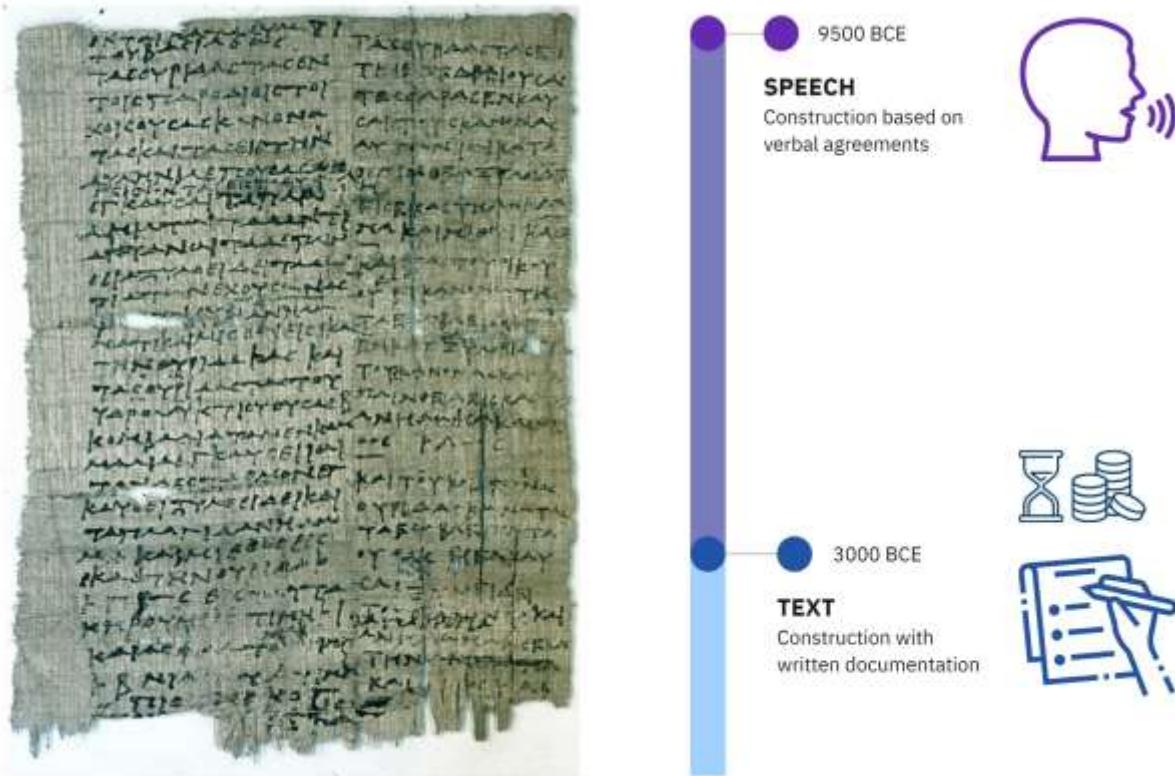


Fig. 1.1-2 En papyrus fra 300-tallet f.Kr. som beskriver kostnadene ved å male ulike typer vinduer i et kongelig palass ved hjelp av enkaustisk teknikk.

Det tok menneskeheden omrent 5000 år å gå fra muntlige samtaler til skriftlige dokumenter i bygg- og anleggsprosjekter, og like lang tid tok det å gå fra papir til digitale data som den primære ressursen for planlegging og kontroll.

På samme måte som utviklingen av handel og pengeforhold stimulerte fremveksten av skriftspråket og de første advokatene for å løse tvister, førte de første registreringene av materialkostnader og arbeidsomfang i byggebransjen til fremveksten av de første lederne i byggebransjen, som blant annet hadde som oppgave å dokumentere, overvåke og være ansvarlig for nøkkelinformasjon om prosjektets tidslinjer og kostnader.

I dag spiller data en mye viktigere rolle: De registrerer ikke bare de beslutningene som er tatt, men blir også et verktøy for å forutsi og modellere fremtiden. Dette er grunnlaget for den moderne prosesstilnærmingen i prosjektledelse - å gjøre akkumulert erfaring om til et beslutningssystem basert på strukturerede og verifiserbare data.

Prosess som verktøy for datadrevne opplevelser

Kjernen i enhver prosess er å omdanne tidligere erfaringer til et verktøy for å planlegge fremtiden. Erfaring i moderne forstand er et strukturert sett med data som kan analyseres for å komme med velbegrunnede spådommer.

Det er historiske data som danner grunnlaget for prognosene, ettersom de tydelig viser resultatene av det arbeidet som er utført, og gir innsikt i hvilke faktorer som påvirker disse resultatene.

La oss ta et konkret eksempel fra monolittisk konstruksjon: Når man planlegger tidspunktet for arbeidene, tar man vanligvis hensyn til betongvolumet, konstruksjonens kompleksitet og værforholdene. La oss anta at en bestemt formann på byggeplassen eller selskapets historiske data for de siste tre årene (2023-2025) viser at det tok mellom 4,5 og 6 dager å støpe en 200 m² stor monolittisk konstruksjon i regnvær (fig. 1.1-3). Det er denne akkumulerte statistikken som danner grunnlaget for å forutsi ledetider og kostnadsressurser ved planlegging av lignende arbeid i fremtidige prosjekter. På grunnlag av disse historiske dataene kan arbeidslederen eller kalkulatøren lage en erfaringsbasert prognose for hvor lang tid det vil ta å fullføre lignende arbeider i 2026 under lignende forhold.

I dette tilfellet med tidsanalyse fungerer analyseprosessen som en mekanisme for å omdanne ulike data til strukturerte opplevelser og deretter til et presist planleggingsverktøy. Data og prosesser er ett økosystem der det ene ikke kan eksistere uten det andre.

Tell det som kan telles, mål det som kan måles, og gjør det som ikke kan måles, målbart [8].

- Galileo Galilei

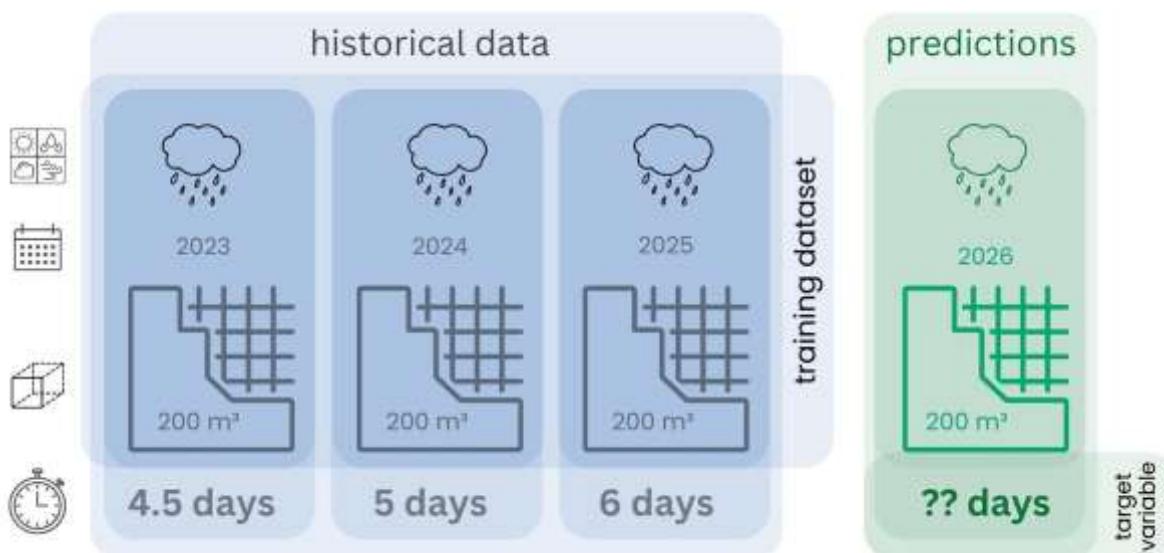


Fig. 1.1-3 Historiske data fungerer som et treningsdatasett for å forutsi en av verdiene i fremtiden

I dagens forretningslandskap er dataanalyse i ferd med å bli en kritisk komponent i effektiv prosjektstyring, prosessoptimalisering og strategisk beslutningstaking. Byggebransjen er gradvis i ferd med å beherske fire viktige analysenivåer, som hver for seg besvarer et spesifikt spørsmål og gir unike fordeler (fig. 1.1-4):

- **Deskriktiv analyse** - besvarer spørsmålet "hva skjedde?" og gir historiske data og rapporter om tidligere hendelser og resultater: I løpet av de siste tre årene (2023-2025) har det tatt mellom 4,5 og 6 dager å støpe en 200 m² stor monolittisk konstruksjon i regnvær.
- **Diagnostisk analyse** - besvarer spørsmålet "hvorfor skjedde dette?" ved å identifisere årsakene til problemene: analysen viser at støpetiden for den monolittiske konstruksjonen økte på grunn av regnvær, noe som bremset herdingsprosessen av betongen
- **Prediktiv analyse** - fremtidsrettet, forutser mulige risikoer og ledetider ved å svare på spørsmålet "hva kommer til å skje?": Basert på historiske data forutsies det at det vil ta omtrent 5,5 dager å støpe en lignende 200 m² stor monolittisk konstruksjon i regnvær i 2026, når man tar hensyn til alle kjente faktorer og trender.
- **Prescriptive analytics** - gir automatiserte anbefalinger og svarer på spørsmålet "hva skal jeg gjøre?", slik at bedriftene kan velge de optimale tiltakene: For å optimalisere arbeidet anbefales det for eksempel å: bruke spesielle tilsetningsstoffer for å fremskynde herdingen av betongen ved høy luftfuktighet, planlegge støping i perioder med lavest sannsynlighet for nedbør og sørge for midlertidige tilfluktsrom for konstruksjonen, noe som vil redusere arbeidstiden til 4-4,5 dager selv under ugunstige værforhold.

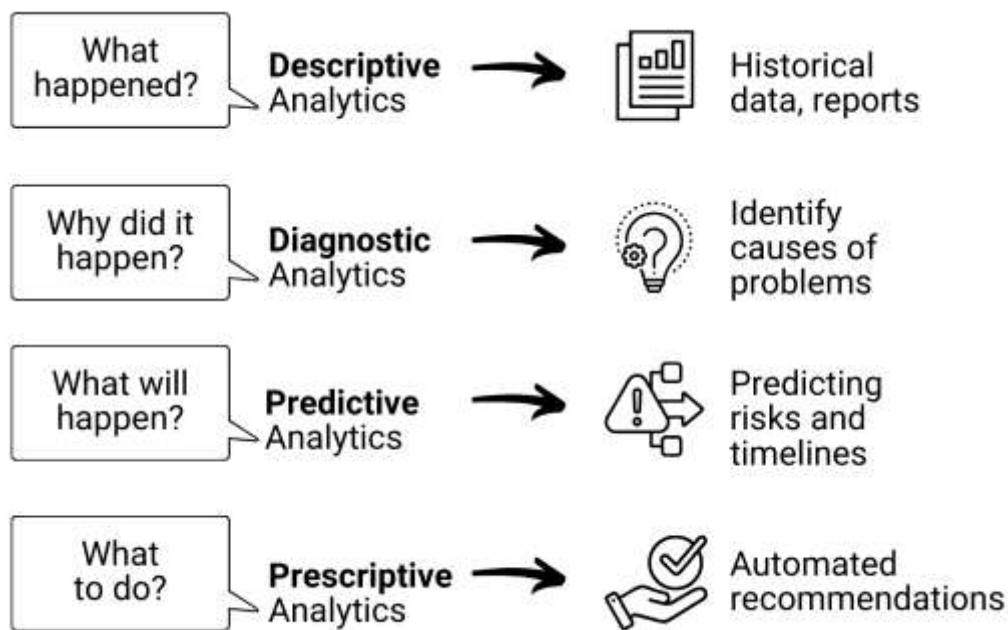


Fig. 1.1-4 Hovedtypene av analyser: fra tidligere beskrivelse til automatisert beslutningstaking.

En fullverdig digital transformasjon, som innebærer en overgang til systemanalyse og datadrevet styring, krever ikke bare outsourcing, men også at det opprettes et kompetent internt team. De viktigste medlemmene i et slikt team bør være produktsjefer, dataingeniører, analyticere og utviklere, som skal jobbe i tett samarbeid med forretningsenhetene (fig. 4.3-9). Dette samarbeidet er nødvendig for å kunne stille intelligente analysespørsmål og effektivt parametrisere bedriftens beslutningsoppgaver. I et informasjonssamfunn blir data ikke bare et hjelpeverktøy, men grunnlaget for prognosering og optimalisering.

I byggebransjen er den digitale transformasjonen i ferd med å endre måten anlegg utformes, forvaltes og drives på. Denne prosessen omtales som digitalisering av informasjon - der alle aspekter av byggeprosessen digitaliseres til en digital form som er egnet for analyse.

Digitalisering av informasjon om byggeprosessen

I årtusener har mengden informasjon som er registrert i bygg og anlegg knapt endret seg, men den har vokst raskt de siste tiårene (figur 1.1-5).

Ifølge PwC-studien® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], ble 90% av alle data i verden skapt i løpet av de siste to årene (per 2015). De fleste bedrifter utnytter imidlertid ikke disse dataene fullt ut, ettersom de enten blir liggende i silosystemer eller bare arkiveres uten reell analyse.

Økningen i datavolumet har bare akselerert de siste årene, med en fordobling fra 15 zettabytes i 2015 til 181 zettabytes i 2025 [10]. Hver dag fylles serverne til bygg- og designfirmaer med prosjektdokumentasjon, arbeidsplaner, beregninger, kalkyler og økonomiske rapporter. For 2D/3D-tegninger brukes formatene DWG, DXF og DGN, og for 3D-modeller - RVT, NWC, PLN og IFC™. Tekstdokumenter, tabeller og presentasjoner

lagres i DOC, XLSX og PPT. I tillegg til video og bilder fra byggeplassen - i MPG og JPEG, sanntidsdata fra IoT-komponenter, RFID® tags (identifikasjon og sporing) og BMS bygningsstyringssystemer (overvåking og kontroll)

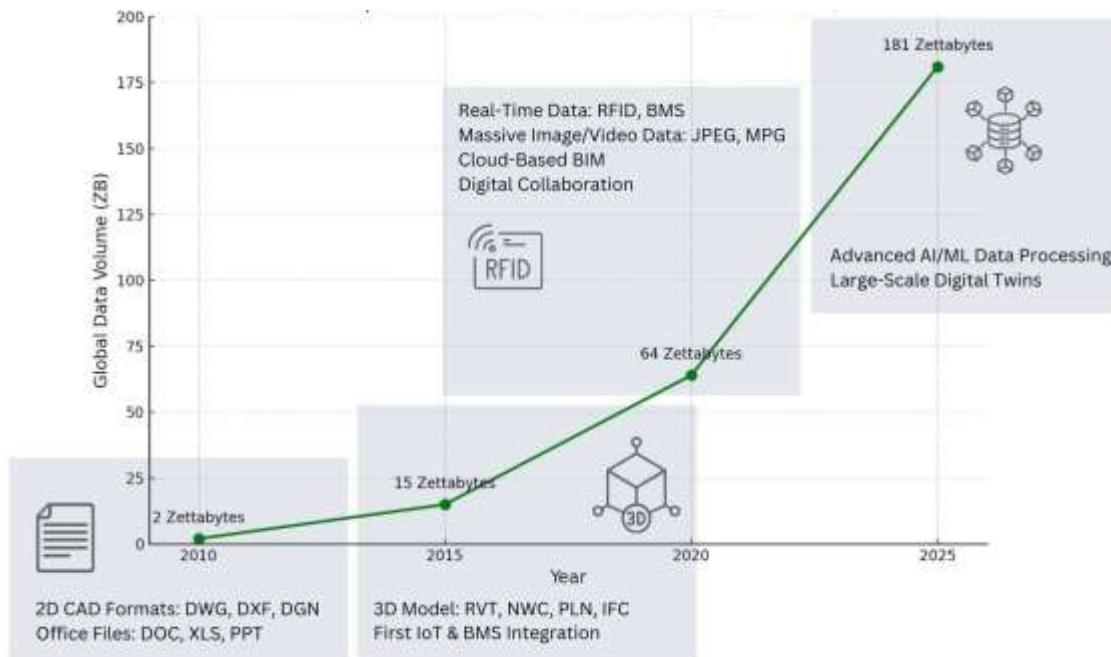


Fig. 1.1-5 Parabolsk datatilvekst 2010-2025 (basert på [10]).

Med den raske informasjonsveksten står byggebransjen overfor et behov for ikke bare å samle inn og lagre data, men også å sikre verifisering, validering, målbarhet og analytisk behandling. I dag er bransjen inne i en aktiv fase av informasjonsdigitalisering - den systematiske omformingen av alle aspekter ved byggevirksomhet til en digital form som eigner seg for analyse, tolkning og automatisering.

Å digitalisere informasjon betyr å ta informasjon om alle enheter og elementer i et byggeprosjekt og selve byggeprosessen - også de vi tidligere ikke betraktet som informasjon i det hele tatt - og konvertere den til et dataformat som gjør informasjonen kvantifiserbar og enkel å analysere.

I forbindelse med bygg og anlegg betyr dette å samle inn og digitalisere informasjon alle elementer i prosjekter og alle prosesser - fra bevegelser av maskiner og mennesker på byggeplassen til vær- og klimaforhold på byggeplassen, aktuelle materialpriser og sentralbankrenter - for å generere analytiske modeller.

Hvis du kan måle det du snakker om, og uttrykke det i tall, vet du noe om emnet. Men hvis du ikke kan uttrykke det kvantitativt, er kunnskapen din ekstremt begrenset og utilfredsstillende. Det kan være et utgangspunkt, men det er ikke nivået for sann vitenskapelig kunnskap. [11].

- W. Thomson (Lord Kelvin), 1824-1907, britisk vitenskapsmann

Digitaliseringen av informasjon går langt utover den tradisjonelle tilnærmingen til informasjonsinnhenting, der man bare registrerte grunnleggende beregninger som arbeidstimer eller faktiske materialkostnader. I dag kan praktisk talt alle hendelser omdannes til en datastrøm som egner seg for dyptgående analyser ved hjelp av avanserte analyseverktøy og maskinlæringsteknikker. Byggebransjen har gjennomgått et fundamentalt skifte fra papirtegninger, Excel-regneark og muntlige instruksjoner til digitale systemer (figur 1.2-4), der hvert eneste element i et prosjekt blir en datakilde. Selv de ansatte - fra ingeniører til bygningsarbeidere på byggeplassen - blir nå sett på som en samling digitale variabler og datasett.

Ifølge KPMGs "Familiar Challenges - New Approaches: Global Construction Survey 2023" er digitale tvillinger, AI (kunstig intelligens) og Big Data i ferd med å bli viktige drivkrefter for å forbedre lønnsomheten i prosjekter [2].

Moderne teknologi forenkler ikke bare informasjonsinnhenting og gjør den i stor grad automatisk, men reduserer også kostnadene ved datalagring radikalt. Resultatet er at bedrifter går bort fra en selektiv tilnærming og foretrekker å lagre hele informasjonsmengden for senere analyse (fig. 2.1-5), noe som åpner for potensielle muligheter for optimalisering av prosesser i fremtiden.

Digitalisering av informasjon og digitalisering gjør det mulig å avdekke skjulte, tidligere uutnyttede verdier i informasjon. Hvis data organiseres på riktig måte, kan de gjenbrukes, tolkes på nytt og integreres i nye tjenester og løsninger.

I fremtiden vil digitaliseringen av informasjon sannsynligvis føre til full automatisering av dokumenthåndtering, innføring av selvstyrte byggeprosesser og fremveksten av nye yrkesgrupper - byggdataanalytikere, AI-prosjektleddere og digitale ingeniører. Byggeprosjekter vil bli dynamiske informasjonskilder, og beslutninger vil ikke være basert på intuisjon eller subjektiv erfaring, men på pålitelige og reproducerte digitale fakta

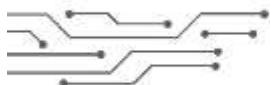
Informasjon er det 21. århundrets olje, og analytics er forbrenningsmotoren [12].

- Peter Sondergaard, Senior Vice President, Gartner®

Ifølge IoT Analytics 2024 [13] forventes det at de globale utgiftene til datahåndtering og -analyse vil øke dramatisk fra 185,5 milliarder dollar i 2023 til 513,3 milliarder dollar innen 2030, med en årlig vekstrate på 16%. Men ikke alle komponentene vokser i samme takt: Analyse vokser raskt, mens veksten innen lagring avtar. Analyse vil stå for den raskeste veksten i økosystemet for datahåndtering: Det forventes å vokse fra 60,6 milliarder dollar i 2023 til 227,9 milliarder dollar innen 2030, en årlig vekstrate på 27%.

Med den akselererende digitaliseringen av informasjon og den raske veksten i informasjonsmengden står ledelsen i byggeprosjekter og -bedrifter overfor et behov for å systematisk lagre, analysere og behandle ulike, ofte heterogene data. Som svar på denne utfordringen begynte bransjen fra midten av 1990-tallet å gå over til elektronisk oppretting, lagring og håndtering av dokumentasjon - fra regneark og designberegninger til tegninger og kontrakter.

Tradisjonelle papirdokumenter, som krever signaturer, fysisk lagring, regelmessig revisjon og arkivering i skap, erstattes gradvis av digitale systemer som lagrer data på en strukturert måte - i databaser med spesialiserte applikasjoner.



KAPITTEL 1.2.

TEKNOLOGIER OG STYRINGSSYSTEMER I MODERNE BYGG OG ANLEGG

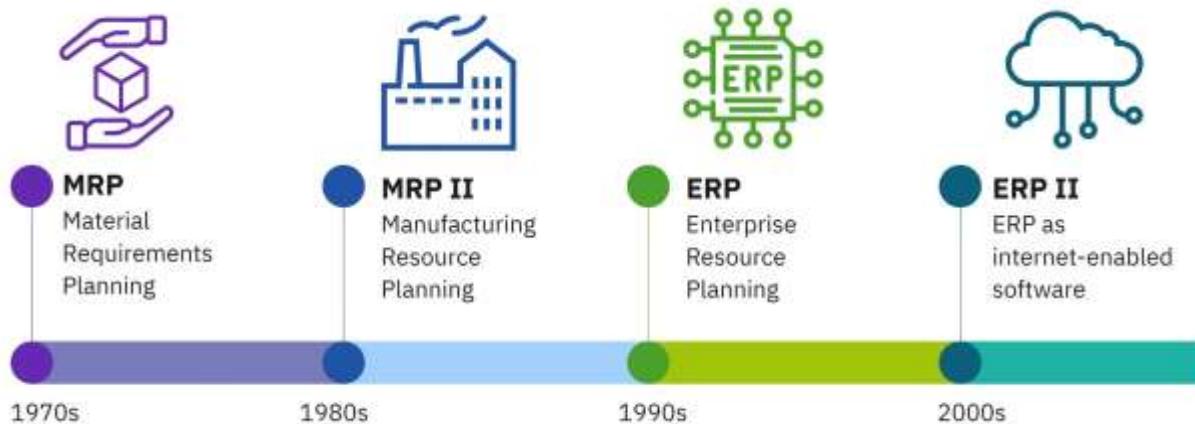
Den digitale revolusjonen og fremveksten av modulære MRP/ERP -systemer

Den moderne digitale datalagringen og -behandlingen begynte med innføringen av magnetbånd på 1950-tallet, som åpnet for muligheten til å lagre og bruke store mengder informasjon. Det neste gjennombruddet var diskettstasjonene, som radikalt endret tilnærmingen til datahåndtering i byggebransjen.

Med utviklingen av datalagring har stort antall selskaper gått inn på markedet for løsninger og begynt å utvikle modulær programvare for å opprette, lagre og behandle data og automatisere rutineoppgaver.

Den eksponentielle veksten av informasjon og verktøy har ført til et behov for integrerte, modulære løsninger som ikke bare fungerer med enkelfiler, men som bidrar til å håndtere og kontrollere dataflyten på tvers av prosesser og prosjekter.

De første omfattende plattformverktøyene skulle ikke bare lagre dokumenter, men også dokumentere alle endringsforespørsler og operasjoner i prosesser: hvem som initierte dem, hva omfanget av forespørslene var, og hva som til slutt ble registrert som en verdi eller et attributt. Til disse formålene trengte man et system som kunne spore nøyaktige beregninger og beslutninger som ble tatt (fig. 1.2-1). Slike plattformer var de første MRP- (Material Requirements Planning) og ERP (Enterprise Resource Planning) -systemene som ble populære fra begynnelsen av 1990-tallet [14]



Figur 1.2-1 Fremskritt innen datalagringsteknologi førte til fremveksten av ERP -systemer på 1980-tallet.

De første MRP - og ERP - systemene la grunnlaget for digitaliseringen av forretningsprosesser og styring av byggeprosjekter. Modulbaserte systemer, som opprinnelig ble utviklet for å automatisere viktige forretningsprosesser, ble etter hvert integrert med flere, mer fleksible og tilpasningsdyktige programvareløsninger.

Disse tilleggsløsningene ble utviklet for databehandling og styring av projektinnhold (fig. 1.2-2), og de

erstattet enten enkelte moduler i store systemer, eller supplerte dem på en effektiv måte ved å utvide funksjonaliteten til hele systemet.

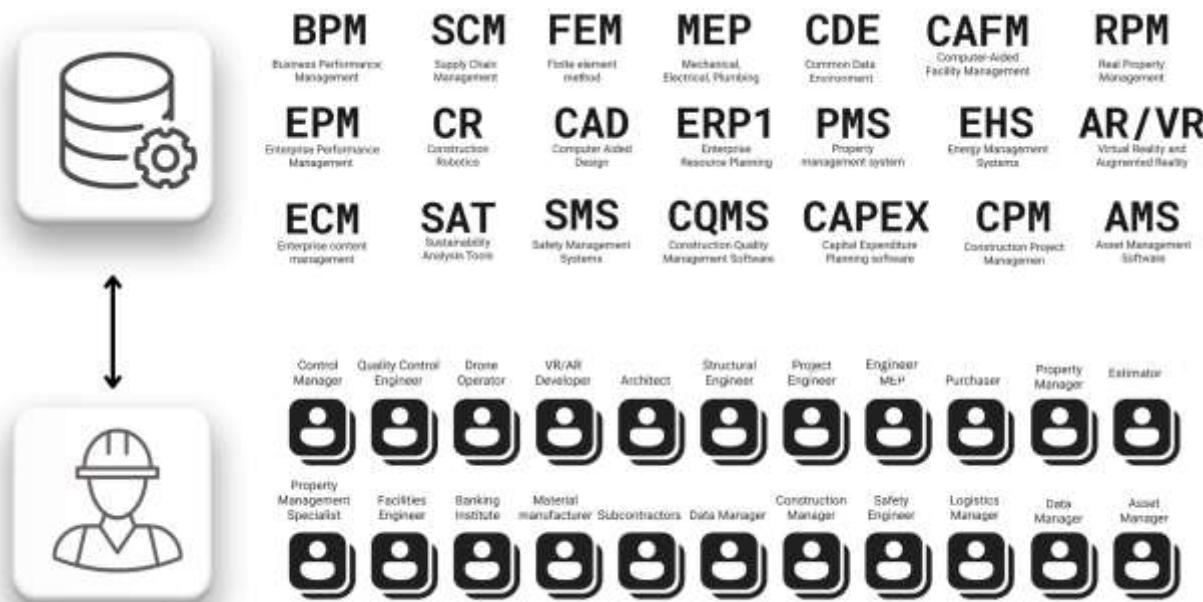


Fig. 1.2-2 Nye programvareløsninger har lokket en hærskare av ledere inn i virksomheten for å administrere dataflyten.

I løpet av de siste tiårene har selskaper investert mye i modulære systemer [15], som de oppfatter som langsiktige, integrerte løsninger.

Ifølge Software Path-rapporten for 2022 [16] er det gjennomsnittlige budsjettet per bruker av et ERP-system på 9 000 dollar. I gjennomsnitt bruker ca. 26% av selskapets ansatte slike systemer. For en organisasjon med 100 brukere vil den totale kostnaden for ERP-implementering dermed komme opp i 900 000 dollar.

Investeringer i proprietære, lukkede, modulære løsninger blir mindre og mindre berettiget på bakgrunn av den raske utviklingen av moderne, fleksible og åpne teknologier. Hvis slike investeringer allerede er gjort, er det viktig å foreta en objektiv revurdering av de eksisterende systemenes rolle: om de fortsatt er nødvendige på lang sikt, eller om funksjonene deres kan revideres og implementeres på en mer effektiv og transparent måte.

Et av hovedproblemene med dagens modulære dataplattformer er at de sentraliserer datahåndteringen i lukkede applikasjoner. Resultatet er at data - en bedrifts viktigste ressurs - blir avhengig av spesifikke programvareløsninger, i stedet for omvendt. Dette begrenser gjenbruk av informasjon, vanskelig gjør migrering og reduserer virksomhetens fleksibilitet i et digitalt landskap i rask endring.

Hvis det er sannsynlig at relevansen eller betydningen av lukket modulær arkitektur vil avta i fremtiden, er det fornuftig å anerkjenne kostnadene som påløper i dag som ugenkallelige kostnader og fokusere på et strategisk skifte til et mer åpent, skalerbart og tilpasningsdyktig digitalt økosystem.

Proprietær programvare kjennetegnes ved at utvikleren har eksklusiv kontroll over kildekoden og brukerdataene som skapes ved bruk av slike løsninger. I motsetning til programvare med åpen kildekode har brukerne ikke tilgang til den interne strukturen i applikasjonen og kan ikke på egen hånd gjennomgå, endre eller tilpasse den til egne behov. I stedet må de kjøpe lisenser som gir rett til å bruke programvaren innenfor de grensene som er fastsatt av leverandøren.

En moderne datasentrisk tilnærming tilbyr et annet paradigme: Data bør betraktes som en strategisk kjerneverdi - uavhengig, holdbar og atskilt fra spesifikke programvareløsninger. Applikasjoner blir i sin tur bare dataverktøy som fritt kan byttes ut uten risiko for å miste kritisk informasjon.

Utviklingen av ERP- og MRP-systemer på 1990-tallet (figur 1.2-1) ga virksomhetene kraftige verktøy for prosesstyring, men hadde også den utilsiktede konsekvensen at antallet personer som var involvert i å opprettholde informasjonsflyten, økte betraktelig. I stedet for å automatisere og forenkle driftsoppgavene skapte disse systemene ofte nye nivåer av kompleksitet, byråkrati og avhengighet av interne IT-ressurser.

Datahåndteringssystemer: fra datautvinning til forretningsutfordringer

Dagens bedrifter står overfor behovet for å integrere flere datahåndteringssystemer. Valg av datahåndteringssystemer, fornuftig styring av disse systemene og integrering av ulike datakilder blir stadig viktigere for virksomhetens resultater.

På midten av 2020-tallet kan du finne hundrevis (og tusenvis i store byggfirmaer) av ulike systemer (figur 1.2-3) som må fungere i harmoni for å få alle aspekter av byggeprosessen til å fungere smidig og sammenhengende.

Ifølge Deloittes studie fra 2016® "Data-Driven Management in Digital Capital Projects" - bruker den gjennomsnittlige bygg- og anleggsarbeideren 3,3 programmer daglig, men bare 1,7 av dem er integrert med hverandre [17].

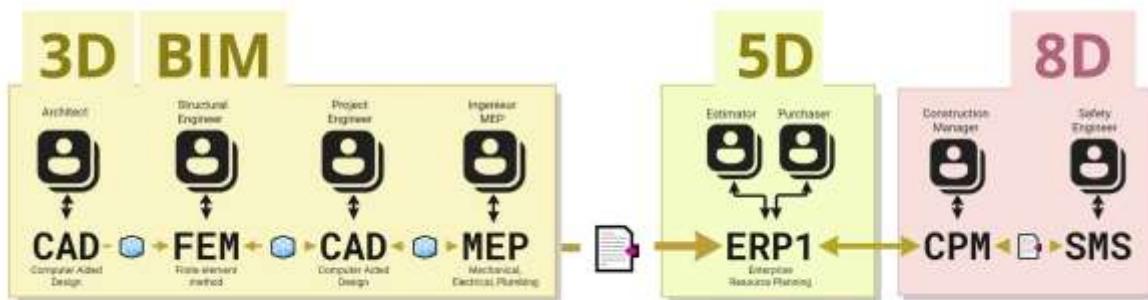


Fig. 1.2-3 Alle forretningsystemer krever et profesjonelt team og en ansvarlig leder for kvalitetsstyring av data.

Her følger en liste over populære systemer som brukes til effektiv styring av byggeprosjekter i mellomstore og store bedrifter i byggebransjen:

- **ERP (Enterprise Resource Planning)** - sørger for integrering av forretningsprosesser, inkludert regnskap, innkjøp og prosjektstyring.
- **CAPEX (Capital Expenditure Planning Software)** - brukes til budsjettering og styring av finansielle investeringer i byggeprosjekter, og hjelper til med å bestemme kostnadene for anleggsmidler og investeringer i langsiktige eiendeler.
- **CAD (Computer-Aided Design) og BIM (Building Information Modeling)** - brukes til å lage detaljerte og nøyaktige tekniske tegninger og 3D-modeller av prosjekter. Disse systemene fokuserer på å arbeide med geometrisk informasjon.
- **MEP (Mechanical, Electrical, Plumbing)** - Ingeniørsystemer som omfatter mekaniske, elektriske og VVS-relaterte komponenter, og som beskriver det interne "sirkulasjonssystemet" i et prosjekt.
- **GIS (geografiske informasjonssystemer)** - brukes til terrenghanalyse og planlegging, inkludert kartografi og romlig analyse.
- **CQMS (programvare for kvalitetsstyring i bygg- og anleggsbransjen)** - sikrer at byggeprosessene er i samsvar med etablerte standarder og forskrifter, noe som bidrar til å eliminere feil og mangler.
- **CPM (Construction Project Management)** - omfatter planlegging, koordinering og kontroll av byggeprosesser.
- **CAFM (Computer-Aided Facility Management)** - systemer for styring og vedlikehold av bygninger.
- **SCM (Supply Chain Management)** er nødvendig for å optimalisere material- og informasjonsflyten mellom leverandørene og byggeplassen.
- **EPM (Enterprise Performance Management)** - har som mål å forbedre forretningsprosesser og resultater.
- **AMS (Asset Management Software)** - brukes til å optimalisere bruk, administrasjon og vedlikehold av utstyr og infrastruktur gjennom hele livssyklusen til eiendelene.
- **RPM (Real Property Management)** - omfatter oppgaver og prosesser knyttet til forvaltning og drift av bygninger og tomter, samt tilhørende ressurser og eiendeler.

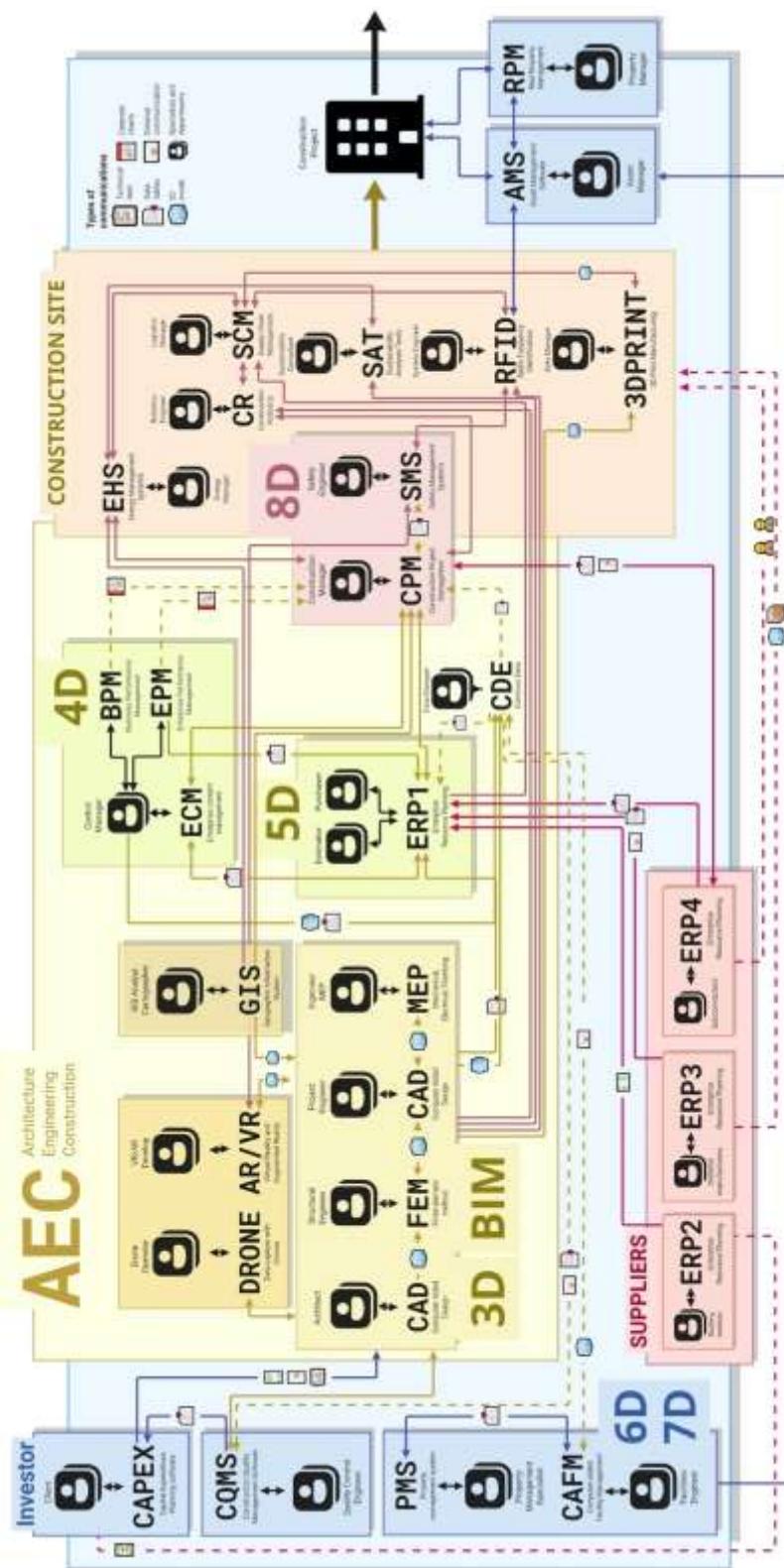


Fig. 1.2-4 Sammenkobling av systemer som knytter bedriftens prosesser sammen med informasjonsflyten mellom ulike avdelinger.

- **CAE (Computer-Aided Engineering)** - Datastøttet ingeniørarbeid, omfatter beregnings- og simuleringssystemer som finite element-analyse (FEA) og CFD (Computational Fluid Dynamics).
- **CFD (Computational Fluid Dynamics)** - Computational fluid dynamics, modellering av væske- og gassstrømmer. CAE-underkategori.
- **CAPP (Computer-Aided Process Planning)** - Datastøttet prosessplanlegging. Det brukes til å lage rute- og prosesskart.
- **CAM (Computer-Aided Manufacturing)** - datastøttet produksjon, generering av styringsprogrammer for CNC-maskiner.
- **PDM (Product Data Management)** - Product Data Management, et system for lagring og håndtering av teknisk dokumentasjon.
- **MES (Manufacturing Execution System)** er et system for styring av produksjonsprosesser i sanntid.
- **PLM (Product Lifecycle Management)** - livssyklusstyring av et prosjektelement, integrerer PDM, CAPP, CAM og andre systemer for fullstendig produktkontroll fra utvikling til avhending.

Disse og mange andre systemer, som omfatter en rekke ulike programvareløsninger, har blitt en integrert del av den moderne byggebransjen (fig. 1.2-4). I bunn og grunn er slike systemer spesialiserte databaser med intuitive grensesnitt som gjør det mulig å legge inn, behandle og analysere informasjon på en effektiv måte i alle faser av prosjekteringen og byggingen. Integreringen av digitale verktøy bidrar ikke bare til å optimalisere arbeidsprosessene, men også til å forbedre nøyaktigheten i beslutningene, noe som har en positiv innvirkning på tidsskjemaet og kvaliteten på prosjektgjennomføringen.

Men i halvparten av tilfellene er det ingen integrasjon. Ifølge statistikken er bare annenhver applikasjon eller hvert annet system integrert med andre løsninger [17]. Dette viser at det digitale miljøet fortsatt er fragmentert, og understreker behovet for å utvikle åpne standarder og enhetlige grensesnitt for å sikre ende-til-ende-informasjonsutveksling i et byggeprosjekt.

En av hovedutfordringene ved integrasjon i moderne bedrifter er fortsatt den høye kompleksiteten i de digitale systemene og kravene til brukerkompetanse som er nødvendig for effektiv informasjonssøking og -tolkning. Et team av spesialister, ledet av en nøkkelansvarlig (fig. 1.2-2), dannes for å støtte hvert system som implementeres i virksomheten.

Den sentrale systemforvalteren spiller en avgjørende rolle for at dataflyten går i riktig retning og er ansvarlig for kvaliteten på den endelige informasjonen, akkurat som de første forvalterne for flere tusen år siden var ansvarlige for tallene som ble skrevet på papyrus eller leirtavler.

For å gjøre ulike informasjonsstrømmer til et styringsverktøy er det avgjørende å kunne integrere og håndtere data på en systematisk måte. I denne arkitekturen må lederne oppstre som elementer i et enkelt nettverk - som et mycel som forbinder de enkelte delene av bedriften til en sammenhengende levende organisme som er i stand til å tilpasse seg og utvikle seg.

Mycelium i bedriften: hvordan data kobles til forretningsprosesser

Prosessen med å integrere data i applikasjoner og databaser er basert på aggregering av informasjon fra en rekke ulike kilder, inkludert ulike avdelinger og spesialister (figur 1.2-4). Spesialister søker etter relevante data, bearbeider dem og overfører dem til sine systemer og applikasjoner for videre bruk.

Hvert, som består av et sett med verktøy, teknologier og databaser, er et kunnskapstre med røtter i jordmonnet av historiske data som vokser og bærer ny frukt i form av ferdige løsninger: dokumenter, beregninger, tabeller, grafer og dashbord (fig. 1.2-5). Systemene i en bedrift, som trærne i et bestemt skogholt, samhandler og kommuniserer med hverandre og utgjør et komplekst og velstrukturert system som støttes og styres av kompetente ledere.

En bedrifts system for informasjonsinnhenting og -overføring fungerer som et komplekst skogsnettverk bestående av trær (systemer) og mycel (forvaltere) som fungerer som ledere og gjenvinnere, og som sørger for at informasjonen overføres og flyter til de riktige systemene. Dette bidrar til å opprettholde en sunn og effektiv flyt og distribusjon av data i selskapet.

Eksparter, som røtter, absorberer rådata i de innledende fasene av et prosjekt og omdanner dem til næringsstoffer for bedriftens økosystem. Data- og innholdshåndteringssystemer (fig. 1.2-4 - ERP, CPM, BIM, etc.) fungerer som kraftige informasjonsmotorveier der denne kunnskapen sirkulerer gjennom alle nivåer i selskapet.

Akkurat som i naturen, der hvert element i økosystemet spiller sin egen rolle, bidrar hver enkelt prosessdeltaker - fraingeniøren til analytikeren - til informasjonsmiljøets vekst og fruktbarhet i en bedrifts forretningslandskap. Disse systemiske "datatrærne" (fig. 1.2-5) er ikke bare mekanismer for innsamling av informasjon, men et konkurransefortrinn som sikrer bedriftens bærekraft.

Skogøkosystemer gjenspeiler overraskende godt hvordan digitale bedriftsstrukturer er organisert. På samme måte som i en skog - fra undervegetasjonen til tretoppene - fordeler virksomhetsstyringen oppgaver til ansvarsnivåer og funksjonelle avdelinger.

Dype og forgrenede trerøtter gir motstandskraft og tilgang til næringsstoffer. På samme måte støtter en solid organisasjonsstruktur og stabile prosesser for arbeid med kvalitetsdata hele informasjonsøkosystemet i et selskap, noe som bidrar til bærekraftig vekst og utvikling selv i perioder med ustabilitet og kriser i markedet.

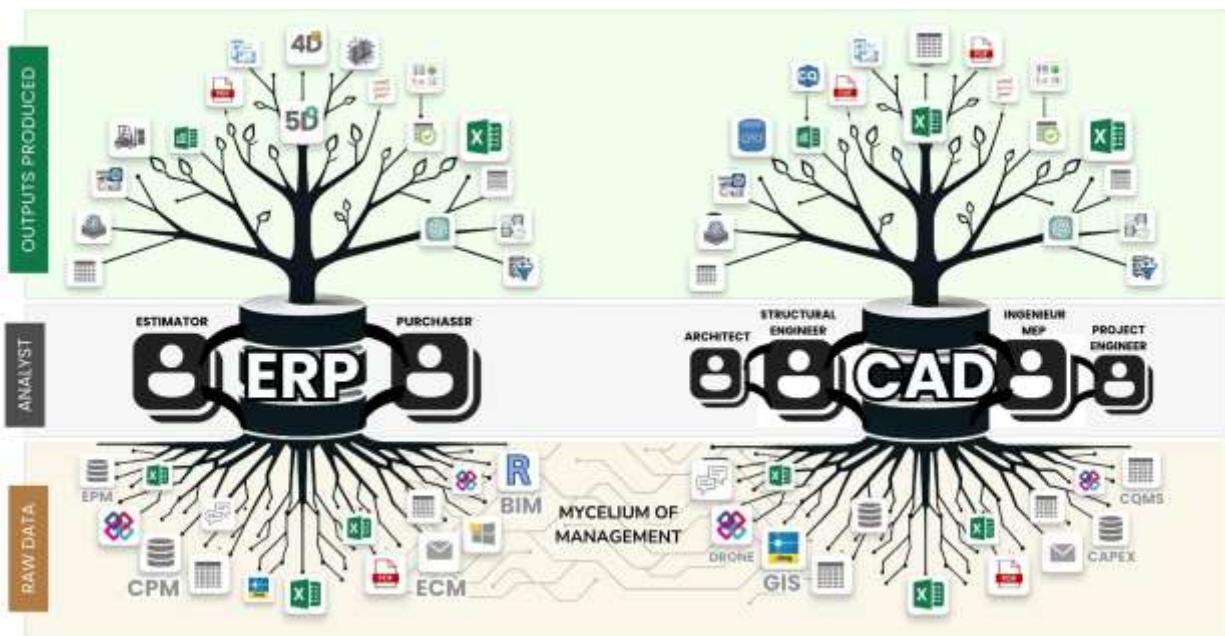


Fig. 1.2-5 Integrering av data gjennom ulike systemer er som et mycelium som knytter ledere og spesialister sammen til ett enkelt informasjonsnettverk.

Den moderne forståelsen av skala i næringslivet har utviklet seg. I dag bestemmes verdien av et selskap ikke bare av den synlige delen - "kronene" i form av slutt-dokumenter og rapporter - men også av dybden i "rotsystemet" av kvalitativt innsamlede og systematisk bearbeidede data. Jo mer informasjon som kan samles inn og bearbeides, desto høyere blir forretnings-verdien. Bedrifter som metodisk samler en "kompost" av allerede bearbeidede data og er i stand til å trekke ut nyttig innsikt fra dem får en strategisk fordel.

Historisk informasjon er i ferd med å bli en ny form for kapital som muliggjør vekst, prosessoptimalisering og konkurransesfortrinn. I en datadrevet verden er det ikke den som har mest, men den som vet mest, som vinner.

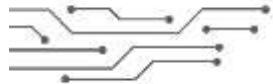
For byggebransjen betyr dette at man går over til prosjektstyring i sanntid, der alle prosesser - fra prosjektering og innkjøp til koordinering av entreprenører - vil være basert på relevante, daglig oppdaterte data. Integrering av informasjon fra ulike kilder (ERP-systemer, CAD-modeller, sensorer IoT på byggeplasser, RFID) vil gjøre det mulig å lage mer nøyaktige prognosenter, reagere raskt på endringer og unngå forsinkelser som skyldes mangel på oppdaterte data.

Ifølge McKinsey & Company's Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]) vil vellykkede selskaper i fremtiden basere seg på data i alle viktige aspekter av virksomheten, fra strategiske beslutninger til operasjonelle interaksjoner.

Data vil ikke lenger bare være et analyseverktøy, men vil bli en integrert del av alle forretningsprosesser og bidra til åpenhet, kontroll og automatisering av ledelsen. Datadrevet på vil gjøre det mulig for organisasjoner å minimere innvirkningen av den menneskelige faktoren, redusere den operasjonelle risikoen og øke gjennomsiktigheten og effektiviteten i beslutningsprosessen.

Det 21. århundret snur opp ned på det økonomiske paradigmet: Mens olje tidligere ble kalt "svart gull" på

grunn av sin evne til å drive maskiner og transport, er historiske data i dag, komprimert under tidspress, i ferd med å bli en ny strategisk ressurs som ikke driver maskiner, men beslutningsalgoritmer som vil drive virksomheten.



KAPITTEL 1.3.

DEN DIGITALE REVOLUSJONEN OG EKSPLOSJONEN AV DATA

Begynnelsen på databoomen som en evolusjonær bølge

Byggebransjen opplever en informasjonseksplosjon uten sidestykke. Hvis vi tenker på virksomheten som et kunnskapstre (fig. 1.2-5) som mates av data, kan den nåværende digitaliseringfasen sammenlignes med den raske veksten av vegetasjon i karboniden, en epoke da jordens biosfære ble forvandlet av den raske akkumuleringen av biomasse (fig. 1.3-1).

Med den globale digitaliseringen fordobles informasjonsmengden i byggebransjen hvert år. Moderne teknologi gjør det mulig å samle inn data i bakgrunnen, analysere dem i sanntid og bruke dem i et omfang som virket umulig for bare kort tid siden.

Ifølge Moores lov, formulert av Gordon Moore (medgrunnlegger av Intel®), fordobles tettheten og kompleksiteten til integrerte kretser og mengden data som behandles og lagres, omtrent annethvert år [19].

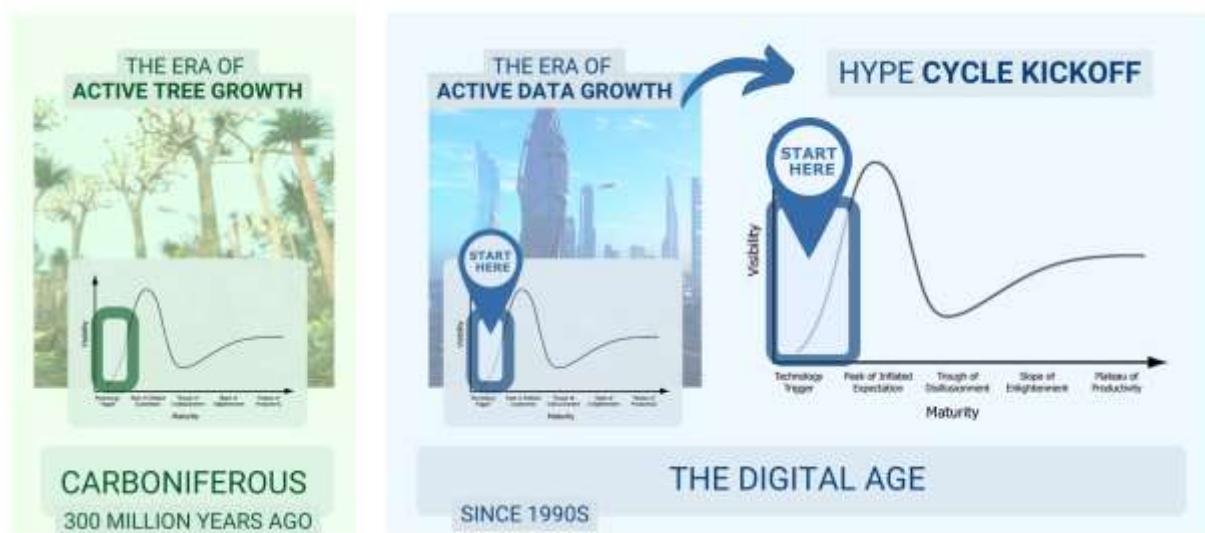


Fig. 1.3-1 Digitaliseringen har ført til en eksponentiell vekst i datamengden, på samme måte som vegetasjonsveksten i kullalderen.

Mens gamle megalittiske strukturer som Göbekli Tepe (Tyrkia) ikke etterlot seg dokumentert kunnskap som var egnet for gjenbruk, gjør digital teknologi det i dag mulig å samle og gjenbruke informasjon. Dette kan sammenlignes med den evolusjonære overgangen fra sporeplanter til frøplanter (angiospermer): Frøets fremvekst ga opphav til den store utbredelsen av liv på planeten. (Fig. 1.3-2).

På samme måte blir data fra tidligere prosjekter en slags "digitale frø" - DNA-kunnskapsbærere som kan skaleres og brukes i nye prosjekter og produkter. Fremveksten av moderne verktøy for kunstig intelligens - maskinlæring og store språkmodeller (LLM-er) som ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN,

Grok - gjør det mulig å hente ut data automatisk, tolke dem og bruke dem i nye sammenhenger.

På samme måte som frø revolusjonerte spredningen av liv på en opprinnelig livløs planet, er "datafrø" i ferd med å bli grunnlaget for den automatiske fremveksten av nye informasjonsstrukturer og kunnskap, slik at digitale økosystemer kan utvikle seg uavhengig og tilpasses seg skiftende brukerbehov.

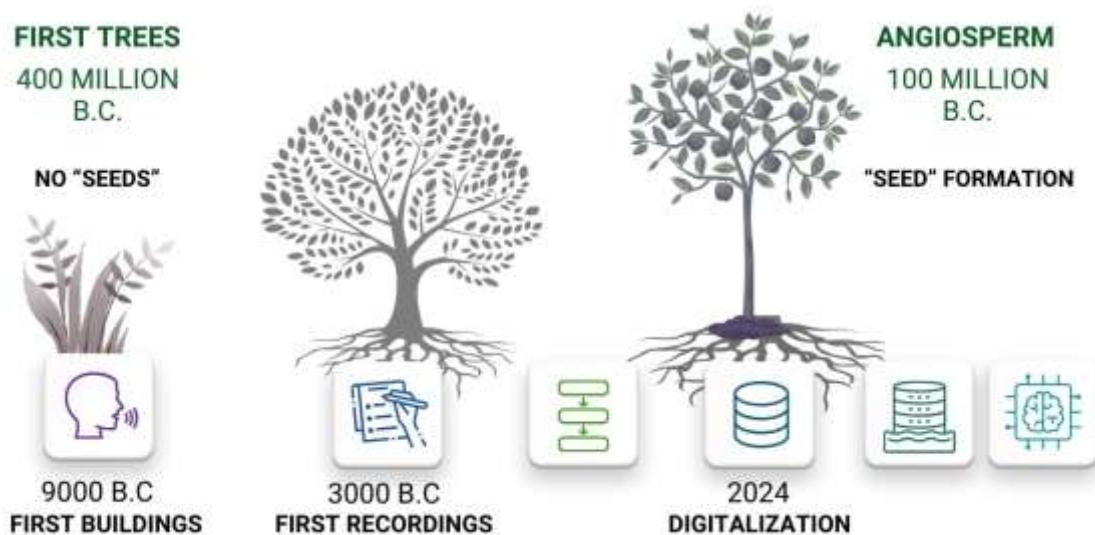


Fig. 1.3-2 Digitale "datafrø" spiller samme evolusjonære rolle som angiospermene, de blomstrende plantene som forandret Jordens økosystem.

Vi står på terskelen til en ny æra i byggebransjen, der eksplosjonen av data og den aktive spredningen av "datafrø" - strukturert informasjon fra tidligere og pågående prosjekter - danner grunnlaget for bransjens digitale fremtid. Ved å "pollinere" disse dataene gjennom store dataspråkmodeller (LLM-er) kan vi ikke bare observere digitale endringer, men aktivt delta i etableringen av selvlærende, adaptive økosystemer. Dette er ikke evolusjon - det er en digital revolusjon der data er i ferd med å bli den viktigste byggestenen i ny virkelighet.

Datamengden i byggebransjen øker dramatisk på grunn av informasjon fra ulike fagområder gjennom hele livssyklusen til byggeprosjekter. Denne enorme opphopningen av data har ført byggebransjen inn i Big Data-æraen [20].

- Prof. Hang Yang, Institutt for byggingeniørvitenskap og arkitektur, Wuhan University of Technology, Wuhan,

Veksten av data i informasjonsalderen minner om evolusjonære prosesser i naturen: Akkurat som utviklingen av skog endret det gamle landskapet på planeten, endrer den nåværende informasjonseksplosjonen landskapet i hele byggebransjen.

Mengden data som genereres i en moderne bedrift

I løpet av de siste to årene har 90% av alle eksisterende data i verden blitt skapt [21]. Fra og med 2023 genererer hver person, inkludert fagfolk i byggebransjen, omtrent 1,7 megabyte data i sekundet [22], og den totale datamengden i verden vil nå 64 zettabyte i 2023 og forventes å overstige 180 zettabyte, eller $180 \cdot 10^{15}$ megabyte, innen 2025 [23].

Denne informasjonseksplosjonen har et historisk forløp - Johannes Gutenbergs oppfinnelse av trykkpressen på 1400-tallet. Bare femti år etter at trykkpressen ble introdusert, ble antallet bøker i Europa fordoblet: I løpet av noen få tiår ble det trykket like mange bøker som det hadde blitt laget for hånd i de foregående 1200 årene [24]. I dag ser vi en enda raskere vekst: Datamengden i verden fordobles hvert tredje år.

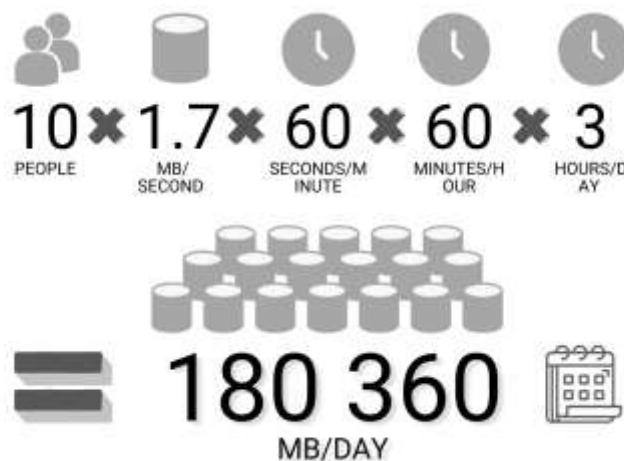
Med dagens datatilvekst har byggebransjen potensial til å generere like mye informasjon i løpet av de neste tiårene som den har samlet inn i hele sin tidligere historie



Fig. 1.3-3 Den daglige lagringen av data fra hver enkelt ansatt på selskapets servere bidrar til en konstant vekst i datamengden.

I dagens byggebransje genererer selv små selskaper en enorm mengde informasjon i flere formater hver eneste dag, og det digitale fotavtrykket til selv et lite byggfirma kan komme opp i flere titalls gigabyte per dag - fra modeller og tegninger til fotografier og sensorer på byggeplassen. Hvis vi antar at hver tekniker i gjennomsnitt genererer ca. 1,7 MB data i sekundet, tilsvarer dette ca. 146 GB per dag, eller 53 TB per år (fig. 1.3-3).

Når et team på 10 personer arbeider aktivt i bare 3 timer daglig, genereres det 180 gigabyte med informasjon per dag (fig. 1.3-4).



Figur 1.3-4 En bedrift med 10 ansatte genererer ca. 50-200 gigabyte med data per dag.

Hvis vi antar at 30% av arbeidsdataene er nye (resten overskrives eller slettes), kan et firma med 10 ansatte skape i størrelsesorden flere hundre gigabyte med nye data i måneden (det faktiske antallet avhenger av hva slags virksomhet bedriften driver med)

Det er dermed klart: Vi genererer ikke bare mer og mer data - vi står også overfor et økende behov for effektiv håndtering, lagring og langsigktig tilgjengelighet. Og mens data tidligere kunne "ligge" på lokale servere uten at det kostet noe, begynner stadig flere bedrifter i forbindelse med den digitale transformasjonen å bruke skyløsninger som grunnlag for informasjonsinfrastrukturen.

Kostnadene ved datalagring: det økonomiske aspektet

De siste årene har stadig flere selskaper outsourset datalagring til skytjenester. Hvis et selskap for eksempel lagrer halvparten av dataene sine i skyen til en gjennomsnittspris på 0,015 dollar per gigabyte per måned, kan lagringskostnadene øke med 10-50 dollar [25] hver måned.

For en liten bedrift med typiske datagenereringsmønstre kan kostnadene for skylagring variere fra noen hundrelapper til potensielt over tusen dollar per måned (figur 1.3-5) i løpet av noen få år, noe som kan utgjøre en betydelig økonomisk byrde.

Ifølge Forresters studie "Enterprises Outsource Data Storage as Complexity Grows" [26], som ble gjennomført blant 214 beslutningstakere innen teknologiinfrastruktur [26], som omfattet en undersøkelse blant 214 beslutningstakere innen teknologiinfrastruktur, outsourcer mer enn en tredjedel av organisasjonene lagring for å håndtere det økende volumet og kompleksiteten i datadriften, og nesten to tredjedeler av virksomhetene foretrekker en abonnementsbasert modell.

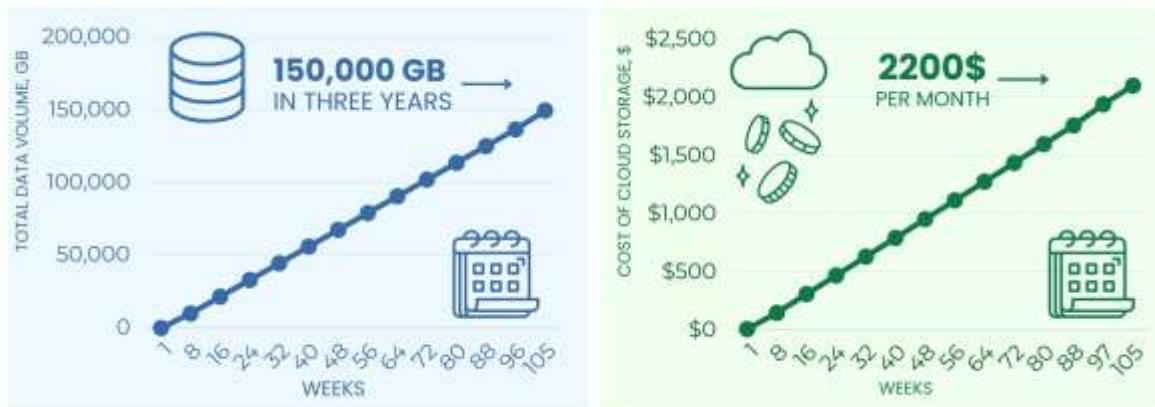


Fig. 1.3-5 Flytting av data til skyen kan øke de månedlige lagringskostnadene med opptil 2 000 dollar, selv for en bedrift med bare 10 ansatte.

Situasjonen kompliseres ytterligere av den akselererende innføringen av skybaserte teknologier som CAD (BIM), CAFM, PMIS og ERP - systemer som øker datalagrings- og prosesseringskostnadene ytterligere. Som følge av dette er bedriftene tvunget til å se seg om etter måter å optimalisere kostnadene på og redusere avhengigheten av skyleverandører.

Siden 2023, med den aktive utviklingen av store språkmodeller (LLM), har tilnærmingen til datalagring begynt å endre seg. Stadig flere selskaper vurderer å ta tilbake kontrollen over dataene sine, ettersom det blir tryggere og mer lønnsomt å behandle informasjon på sine egne servere.

I denne sammenhengen er trenden bort fra skybasert lagring og behandling av kun de nødvendige dataene til fordel for lokal distribusjon av LLM og AI -løsninger for bedrifter. Som Microsofts administrerende direktør påpekte i et av sine intervjuer [27], vil AI-agenter styre prosesser i databaser og automatisere funksjonene til ulike systemer i stedet for å være avhengig av flere separate applikasjoner eller skybaserte SaaS-løsninger å utføre ulike oppgaver.

[...] den gamle tilnærmingen til dette [databehandlingsproblemet] var: Hvis du tenker tilbake på hvordan ulike forretningsapplikasjoner håndterte integrasjon, så brukte de konnektorer. Selskapene solgte lisenser for disse konnektorene, og forretningsmodellen var bygget opp rundt dette. SAP [ERP] er et av de klassiske eksemplene: Du fikk bare tilgang til SAP-data hvis du hadde den riktige koblingen. Så det virker for meg som om noe lignende vil dukke opp når det gjelder [AI]-agentinteraksjon [...]. Vår tilnærming, i hvert fall den vi har, er: Jeg tror at konseptet om eksistensen av forretningsapplikasjoner sannsynligvis vil kollapse i en tid med [AI]-agenter. For hvis du tenker på det, er de egentlig databaser med en haug med forretningslogikk

- Satya Nadella, administrerende direktør i Microsoft, intervju med BG2 Channel, 2024. [28]

I dette paradigmet går den datadrevne LLM-tilnærmingen lenger enn klassiske systemer. Kunstig intelligens blir et mellomledd mellom brukeren og dataene (fig. 2.2-3, fig. 2.2-4), noe som eliminerer behovet for flere

mellomliggende grensesnitt og øker effektiviteten i forretningsprosessene. Vi kommer til å snakke mer om denne måten å arbeide med data på i kapittelet "Snu kaos til orden og reduser kompleksiteten".

Mens fremtidens arkitektur fortsatt er i ferd med å ta form, står bedriftene allerede overfor konsekvensene av tidligere beslutninger. De siste tiårenes massive digitalisering, sammen med innføringen av ulike systemer og ukontrollert opphopning av data, har ført til et nytt problem - informasjonsoverbelastning.

Grenser for dataakkumulering: fra masse til mening

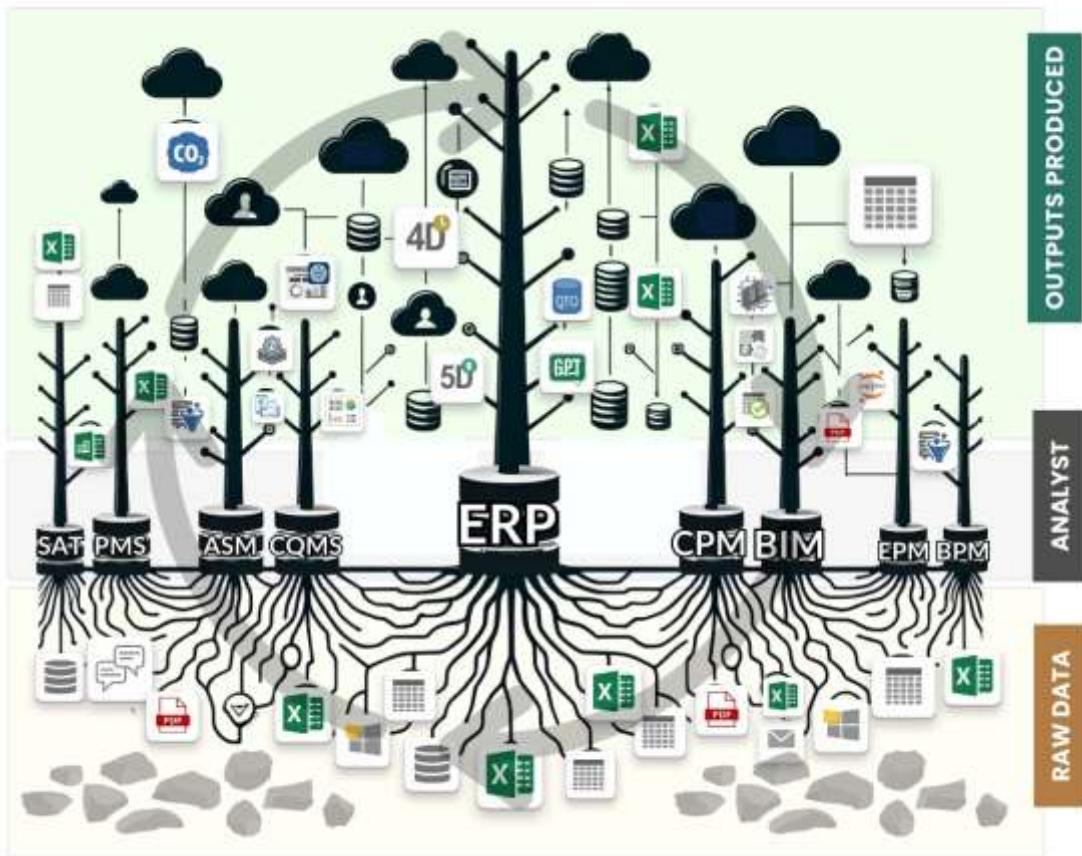
Moderne bedriftssystemer utvikler seg og fungerer under kontrollert vekst, når datamengden og antallet applikasjoner er i balanse med IT-avdelingenes og -ledernes kapasitet. De siste tiårene har imidlertid digitaliseringen ført til en ukontrollerbar økning i datamengden og kompleksiteten, noe som har ført til en overmettingseffekt i bedriftenes informasjonsøkosystem.

I dag utsettes servere og lagringsanlegg for en uhørt stor tilstrømning av ubehandlet og multiformatert informasjon som ikke rekker å bli til kompost og som raskt blir irrelevant. Bedriftens begrensede ressurser kan ikke håndtere denne flommen, og data samler seg i isolerte siloer (såkalte "siloer") som krever manuell behandling for å trekke ut nyttig informasjon.

Som et resultat av dette lider moderne bedriftsstyringssystemer ofte av informasjonsoverbelastning, som en skog overgrodd av eføy og dekket av mugg. I stedet for å gi næring til informasjonshumus i hjertet av bedriftens økosystem, dannes det isolerte områder med data i ulike formater, noe som uunngåelig fører til en reduksjon i forretningsprosessenes samlede effektivitet.

Den lange perioden med eksponentiell datatilvekst de siste 40 årene vil uunngåelig bli etterfulgt av en metningsfase og en påfølgende nedkjølingsfase. Når lagringsmulighetene når sine grenser, vil det skje et kvalitativt skifte: Data vil ikke lenger bare være et lagringsobjekt, men en strategisk ressurs.

Med utviklingen av kunstig intelligens og maskinlæring har selskapene mulighet til å redusere kostnadene ved informasjonsbehandling og gå fra kvantitativ vekst til kvalitativ bruk av data. I løpet av det neste tiåret må byggebransjen skifte fokus fra å skape mer og mer data til å sikre dataenes struktur, integritet og analytiske verdi.



Figur 1.3-6 Isolerte datakilder hindrer informasjonsdeling mellom datasystemer.

Hovedverdien ligger ikke lenger i mengden informasjon, men i evnen til å tolke den automatisk og omdanne den til anvendt kunnskap som er nyttig for å ta ledelsesbeslutninger. For at data skal bli virkelig nyttige, må de håndteres på riktig måte: samles inn, verifiseres, struktureres, lagres og analyseres i forbindelse med spesifikke forretningsoppgaver.

Dataanalyseprosessen i en bedrift kan sammenlignes med syklusen av liv og forfall for trær i en skog og fremveksten av nye, unge og sterke trær: modne trær dør, brytes ned og blir grobunn for ny vekst. Når prosessene er ferdige, blir de en del av bedriftens informasjonsøkosystem, og til slutt blir de til informasjonshumus som gir næring til den fremtidige veksten av nye systemer og data.

I praksis brytes imidlertid denne syklusen ofte. I stedet for organisk fornyelse oppstår det et lagdelt kaos, som geologiske lag, der nye systemer legges på toppen av gamle uten dypere integrering og strukturering. Resultatet er at det oppstår ulike "siloer" av informasjon, noe som hindrer kunnskapsutveksling og vanskeliggjør datahåndtering.

Neste steg: fra datateori til praktisk endring

Utviklingen av data i byggebransjen er en reise fra leirtavler til moderne modulære plattformer. Utfordringen

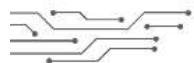
i dag er ikke å samle inn informasjon, men å skape et rammeverk som gjør uensartede og mangfoldige data til en strategisk ressurs. Enten du er bedriftsleder eller ingeniør, vil det å forstå verdien av data og hvordan man arbeider med dem, være en nøkkelkompetanse i fremtiden.

For å oppsummere denne delen er det verdt å fremheve de viktigste praktiske trinnene som vil hjelpe deg med å bruke de diskuterte tilnærmingene i det daglige arbeidet:

- Gjennomfør en personlig revisjon av informasjonsflyten
 - Lag en liste over alle systemene og applikasjonene du jobber med til daglig
 - Merk av hvor du bruker mest tid på å søke eller dobbeltsjekke data
 - Identifiser dine viktigste informasjonskilder
 - Analyser det nåværende applikasjonslandskapet for å avdekke redundans og duplisering av funksjoner
- Streber etter å bevege seg gjennom prosessene etter nivå av analytisk modenhet
 - Start oppgavene dine med beskrivende analyser (hva skjedde?)
 - Innfør gradvis en diagnose (hvorfor skjedde dette?)
 - Tenk på hvordan du kan gå over til prediktiv (hva vil skje?) og preskriptiv (hva skal du gjøre?) analyse i prosessene
- Begynn å strukturere arbeidsdataene dine
 - Implementer et enhetlig system for navngivning av filer og mapper som du ofte bruker i arbeidet ditt
 - Lag maler for dokumenter og rapporter som brukes ofte
 - Arkiver regelmessig fullførte prosjekter med en klar struktur

Selv om du ikke kan endre hele informasjonsinfrastrukturen i teamet eller bedriften din, kan du begynne med dine egne prosesser og små forbedringer i det daglige arbeidet. Husk at den virkelige verdien av data ikke ligger i mengden, men i evnen til å trekke ut nyttig innsikt fra dem. Selv små, men riktig strukturerte og analyserte datasett kan ha en betydelig innvirkning når de integreres i beslutningsprosesser.

I de neste delene av boken går vi videre til spesifikke metoder og verktøy for arbeid med data, ser på måter å omdanne ustrukturert informasjon til strukturerte sett, utforsker teknologier for automatisering av analyser og beskriver hvordan man bygger opp et effektivt økosystem for analyse i et byggefirma.





II DEL

HVORDAN BYGGEBRANSJEN DRUKNER I DATAKAOS

Den andre delen er viet en kritisk analyse av utfordringene byggefirmaer står overfor når de skal håndtere økende datamengder. Konsekvensene av informasjonsfragmentering og fenomenet "data i siloer", som hindrer effektiv beslutningstaking, blir undersøkt i detalj. Problemene med HiPPO -tilnærmingen (Highest Paid Person's Opinion) og dens innvirkning på kvaliteten på ledelsesbeslutninger i byggeprosjekter blir undersøkt. Vi ser på hvordan dynamiske forretningsprosesser og deres økende kompleksitet påvirker informasjonsflyten og effektiviteten i driften. Det gis konkrete eksempler på hvordan overdreven systemkompleksitet øker kostnaden og reduserer organisasjoners fleksibilitet. Det legges særlig vekt på begrensningene som proprietære formater skaper, og utsiktene for bruk av åpne standarder i byggebransjen. Konseptet med å gå over til AI- og LLM-baserte programvareøkosystemer presenteres, som minimerer overdreven kompleksitet og tekniske barriérer.

KAPITTEL 2.1.

DATAFRAGMENTERING OG SILOER

Jo flere verktøy, desto mer effektiv blir virksomheten?

Ved første øyekast kan det virke som om flere digitale verktøy fører til større effektivitet. I praksis er dette imidlertid ikke tilfelle. For hver ny løsning, det være seg en skytjeneste, et eldre system eller en ny Excel-rapport, legger bedriften til enda et lag i det digitale landskapet - et lag som ofte ikke er integrert med resten (figur 2.1-1).

Data kan sammenlignes med kull eller olje: Det tar årevis å bygge dem opp, komprimert under lag av kaos, feil, ustrukturerte prosesser og glemte formater. For å hente ut virkelig nyttig informasjon må bedriftene bokstavelig talt vasse gjennom lag av utdaterte løsninger og digital støy.

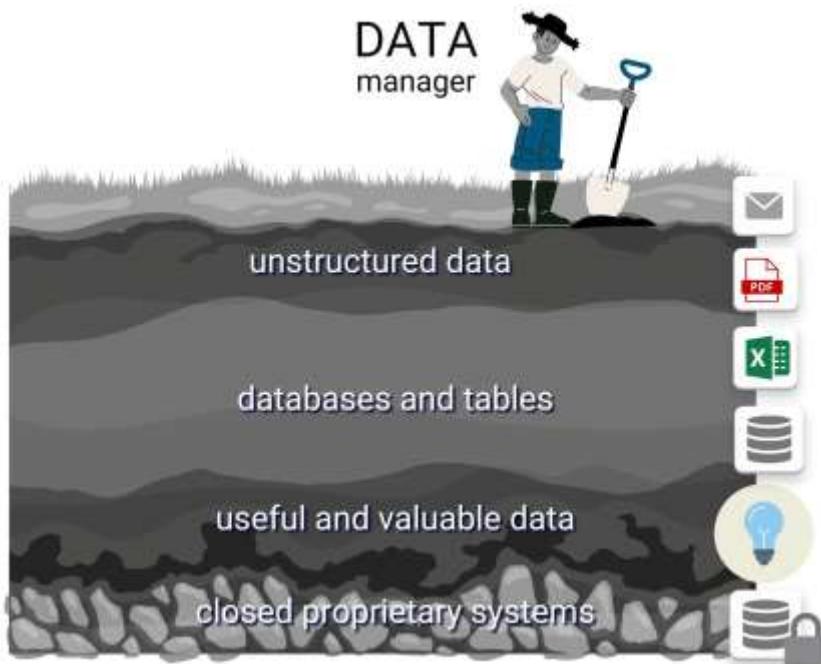


Fig. 2.1-1 Ulike data danner lagdelte lag - selv "gylne" innsikter går tapt i geologiske bergarter av systemisk kompleksitet.

Hver ny applikasjon etterlater seg et spor: en fil, en tabell eller en hel isolert "silo" på serveren. Ett lag er leire (utdaterte og glemte data), et annet er sand (ulike tabeller og rapporter), og det tredje er granitt (lukkede, proprietære formater som ikke kan integreres). Over tid ligner en bedrifts digitale miljø mer og mer på et reservoar av ukontrollert opphopning av informasjon, der verdien går tapt dypt inne i bedriftens servere.

For hvert nye prosjekt og hvert nye system blir ikke bare infrastrukturen mer kompleks, men også veien til nyttige kvalitetsdata. For å komme frem til den verdifulle "steinen" kreves det dyp rensing, strukturering av informasjon, "chunking", gruppering i meningsfulle biter, og utvinning av strategisk viktig innsikt gjennom analyse og datamodellering.

Data er verdifulle, og de vil være lenger enn selve systemene [som behandler dataene] 29].

- Tim Berners-Lee, World Wide Webs far og skaperen av det første nettstedet

Før data kan bli en "verdifulk ting" og et pålitelig beslutningsgrunnlag, må de gjennom en grundig bearbeiding. Det er riktig forbehandling som gjør uensartede data om til strukturert erfaring, nyttig informasjonshumus, som deretter blir et verktøy for prognosering og optimalisering.

Det er en misforståelse at man må ha helt rene data for å begynne å analysere, men i praksis er det å kunne jobbe med skitne data en viktig del av prosessen.

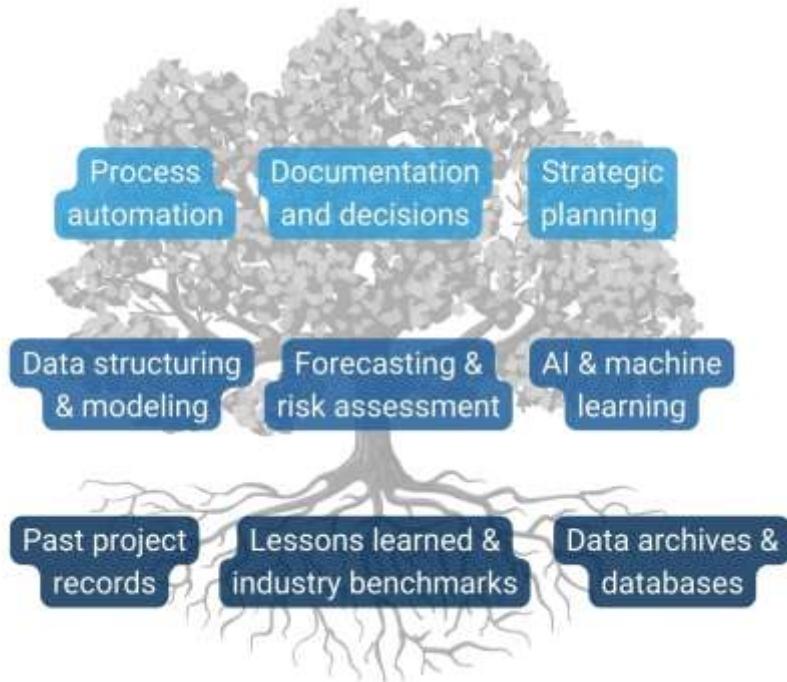


Fig. 2.1-2 Data er rotsystemet og grunnlaget for virksomheten, som igjen er basert på beslutningsprosesser.

I takt med den teknologiske utviklingen må også virksomheten din utvikle seg videre og lære seg å skape verdi av data. På samme måte som olje- og kullselskaper bygger opp infrastrukturen for å utvinne mineraler, må også bedrifter lære seg å håndtere flyten av ny informasjon på sine egne servere og trekke ut verdifull innsikt fra ubrukete, uformaterde og utdaterte data, slik at de kan bli en strategisk ressurs.

Å opprette felt (datavarehus) er det første steget. Selv de kraftigste verktøyene løser ikke problemet med dataisolering og data i flere formater hvis bedriftene fortsetter å operere i silosystemer. Når data eksisterer adskilt fra hverandre, uten at de krysser hverandre og deler informasjon, oppstår det en "datasilo"-effekt. I

stedet for å ha én enkelt, konsistent infrastruktur tvinges bedriftene til å bruke ressurser på å slå sammen og synkronisere data.

Datasiloer og deres innvirkning på selskapets resultater

Tenk deg at du bygger et boligfelt, men at hvert team har sitt eget prosjekt. Noen bygger veger, andre legger kommunikasjon, og andre igjen legger veier uten å sjekke med hverandre. Resultatet er at rørene ikke passer med åpningene i veggene, heissjaktene passer ikke med etasjene, og veiene må demonteres og legges på nytt.

Denne situasjonen er ikke bare et hypotetisk scenario, men en realitet i mange moderne byggeprosjekter. På grunn av det store antallet hoved- og underentreprenører som arbeider med ulike systemer og uten ett koordinerende senter, blir prosessen til en serie med endeløse godkjenninger, omarbeidinger og konflikter. Alt dette fører til betydelige forsinkelser og økte prosjektkostnader.

En klassisk situasjon på en byggeplass er enkel: Forskalingen er klar, men leveransen av armering har ikke kommet i tide. Når man sjekker informasjonen i ulike systemer, er kommunikasjonen omrent som følger:

- ⌚ **Arbeidslederen** på byggeplassen den 20. skriver til prosjektlederen: "Vi er ferdige med forskalingen, men hvor er armeringsjernet?"
- ⌚ **Prosjektleder** (PMIS) til innkjøpsavdelingen: - "Forskalingen er klar. I systemet mitt [PMIS] står det at armeringsjernet skulle komme den 18. Hvor er armeringsjernet?"
- ⌚ **Leverandørkjedespesialist** (ERP): - "ERP-systemet vårt sier at leveransen vil skje den 25."
- ⌚ **Dataingeniør** eller IT-avdelingen (ansvarlig for integrasjoner): - I PMIS er datoan den 18., i ERP er den 25. Det finnes ingen OrderID-kobling mellom ERP og PMIS, så dataene er ikke synkronisert. Dette er et typisk eksempel på et informasjonsgap.
- ⌚ **Prosjektleder** til daglig leder - "Leveransen av innredning er forsinket, byggeplassen står stille, og det er uklart hvem som har ansvaret".

Årsaken til hendelsen var isolasjonen av data i ulike systemer. Ved å integrere og forene datakilder, opprette et felles informasjonslager og automatisere ved hjelp av ETL -verktøy (Apache NiFi, Airflow eller n8n), kan siloene mellom systemene elimineres. Disse og andre metoder og verktøy vil bli diskutert i detalj i senere deler av boken.

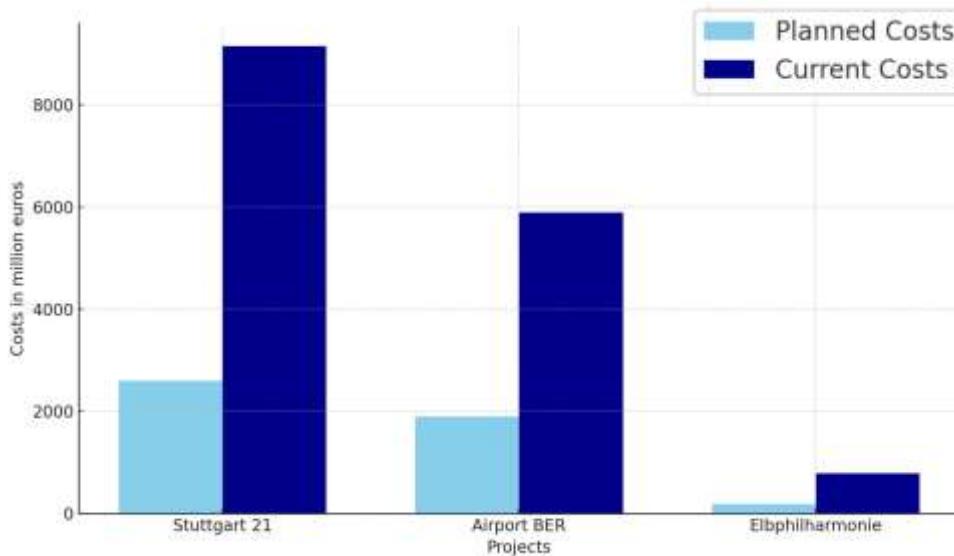


Fig. 2.1-3 Sammenligning av planlagte og faktiske kostnader for store infrastrukturprosjekter i Tyskland.

Det samme skjer med virksomhetssystemer: Først lager man isolerte løsninger, og så må man bruke enorme budsjetter på å integrere og harmonisere dem. Hvis data- og kommunikasjonsmodellene hadde vært gjennomtenkt fra starten av, hadde det ikke vært behov for integrasjon i det hele tatt. Silodata skaper kaos i den digitale verden, som en ukoordinert byggeprosess.

Ifølge KPMGs studie "Cue construction 4.0: Time to make or break" fra 2023 deler bare 36% av bedriftene data effektivt på tvers av avdelinger, mens 61% står overfor alvorlige problemer på grunn av isolerte "datasiloer" [30].

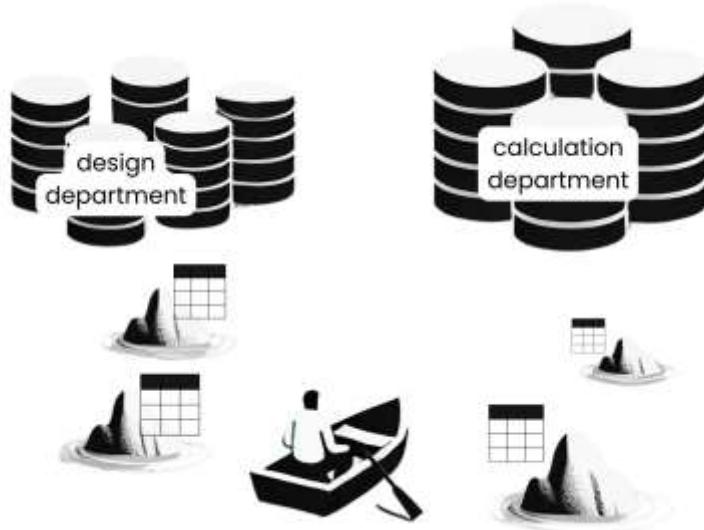


Fig. 2.1-4 År med data som er vanskelig å samle inn, hoper seg opp i isolerte "siloer" med fare for aldri å bli brukt.

Bedriftsdata lagres i isolerte systemer, som enkeltstående trær spredt utover landskapet. Hvert av dem

inneholder verdifull informasjon, men mangelen på forbindelser mellom dem gjør det umulig å skape et enkelt, sammenkoblet økosystem. Denne siloen hindrer dataflyten og begrenser organisasjonens evne til å se hele bildet. Å koble disse siloene sammen er en ekstremt lang og kompleks prosess der man dyrker soppmycel på ledelsesnivå for å lære seg hvordan man overfører enkeltopplysninger mellom systemene.

Ifølge en WEF-studie fra 2016 er mangelen på felles datastandarder og fragmentering en av de største hindringene for digital transformasjon.

Byggebransjen er en av de mest fragmenterte i verden, og den er avhengig av at alle aktørene i verdikjeden samhandler på en smidig måte [5].

- World Economic Forum 2016: Fremtidens

Designere, ledere, koordinatorer og utviklere foretrekker ofte å jobbe selvstendig og unngå kompleksiteten ved koordinering. Denne naturlige tilbøyeligheten fører til at det opprettes "siloer" av informasjon, der data isoleres i separate systemer. Jo flere slike isolerte systemer det finnes, desto vanskeligere er det å få dem til å fungere sammen. Etter hvert får hvert system sin egen database og en spesialisert støtteavdeling med ledere (figur 1.2-4), noe som gjør integrasjonen enda mer komplisert.

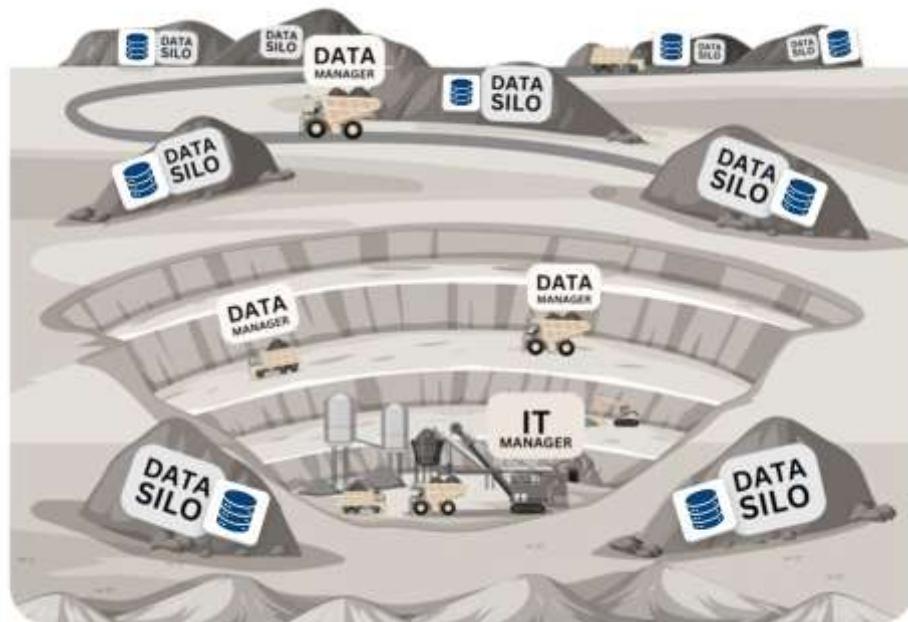


Fig. 2.1-5 Hvert system har en tendens til å skape sin egen silo med data som må behandles av egnede verktøy [31].

Den onde sirkelen i bedriftens systemer ser slik ut: Bedriftene investerer i komplekse løsninger, står overfor høye integrasjonskostnader, og utviklerne, som innser hvor komplisert det er å kombinere systemer, foretrekker å arbeide i sine lukkede økosystemer. Alt dette øker fragmenteringen av IT-landskapet og gjør det vanskeligere å migrere til nye løsninger (figur 2.1-5). Ledere ender opp med å kritisere datasiloer, men analyserer sjeldent årsakene til dem og hvordan de kan forebygges. Ledere klager over utdaterte IT-

systemer, men å bytte dem ut krever betydelige investeringer og gir sjeldent de forventede resultatene. Resultatet er at selv forsøk på å bekjempe problemet ofte gjør vondt verre.

Hovedårsaken til denne avkoblingen er at man prioriterte applikasjoner fremfor data. Bedrifter utvikler først separate systemer eller kjøper hyllevare fra leverandører, og forsøker deretter å forene dem ved å opprette dupliserte og inkompatible lagrings- og databaser.

For å få bukt med fragmenteringsproblemet kreves det en radikalt ny tilnærming - å prioritere data fremfor applikasjoner. Bedriftene må først utvikle strategier for datahåndtering og datamodeller, og deretter bygge systemer eller kjøpe inn løsninger som fungerer med ett enkelt sett med informasjon i stedet for å skape nye barrierer.

Vi er på vei inn i en ny verden der data kan være viktigere enn programvare.

- Tim O'Reilly, administrerende direktør i O'Reilly Media, Inc.

McKinsey Global Institutes studie "Rethinking Construction: the path to improved productivity" (2016) viser at byggebransjen henger etter andre sektorer når det gjelder digital transformasjon [32]. Ifølge rapporten kan bruk av automatisert datahåndtering og digitale plattformer forbedre produktiviteten betydelig og redusere tap knyttet til inkonsekvente prosesser. Behovet for digital transformasjon understrekkes også i Egan-rapporten (Storbritannia, 1998) [33], som fremhever nøkkelrollen til integrerte prosesser og en samarbeidstilnærming i byggebransjen.

Mens hovedproblemene for dataforvaltere de siste 10 000 årene har vært mangel på data, står brukere og forvaltere nå overfor et problem - en overflod av data som gjør det vanskelig å finne juridisk korrekt informasjon av høy kvalitet.

De ulike datasiloene fører uunngåelig til et alvorlig problem med redusert datakvalitet. Med flere uavhengige systemer kan de samme dataene finnes i ulike versjoner, ofte med motstridende verdier, noe som skaper ytterligere problemer for brukerne som må finne ut hvilken informasjon som er relevant og pålitelig.

Duplicering og manglende datakvalitet som en konsekvens av uenighet

Problemet med datasiloer gjør at ledere må bruke mye tid på å lete etter og avstemme data. For å sikre seg mot kvalitetsproblemer oppretter bedrifter komplekse strukturer for informasjonsstyring der en vertikal av ledere har ansvaret for å søke etter, verifisere og avstemme data. Denne tilnærmingen fører imidlertid bare til økt byråkrati og tregere beslutningsprosesser. Jo mer data det finnes, desto vanskeligere er det å analysere og tolke dem, særlig hvis det ikke finnes noen enhetlig standard for lagring og behandling av dem.

Med den store mengden av applikasjoner og systemer som har vokst som paddehatter det siste tiåret, har problemet med siloer og uhensiktsmessig datakvalitet blitt stadig viktigere for sluttbrukerne. De samme dataene, men med forskjellige verdier, kan nå finnes i forskjellige systemer og applikasjoner (figur 2.1-6). Dette fører til problemer for sluttbrukerne når de skal finne ut hvilken versjon av data som er relevant og korrekt blant de mange tilgjengelige. Dette fører til feil i analyser og i siste instans i beslutningsprosesser.

For å sikre seg mot problemer med å finne de riktige dataene oppretter bedriftsledere et byråkrati av

verifikasjonsansvarlige på flere nivåer. Deres oppgave er å raskt finne, kontrollere og sende de nødvendige dataene i form av tabeller og rapporter, og navigere i labyrinten av ulike systemer.

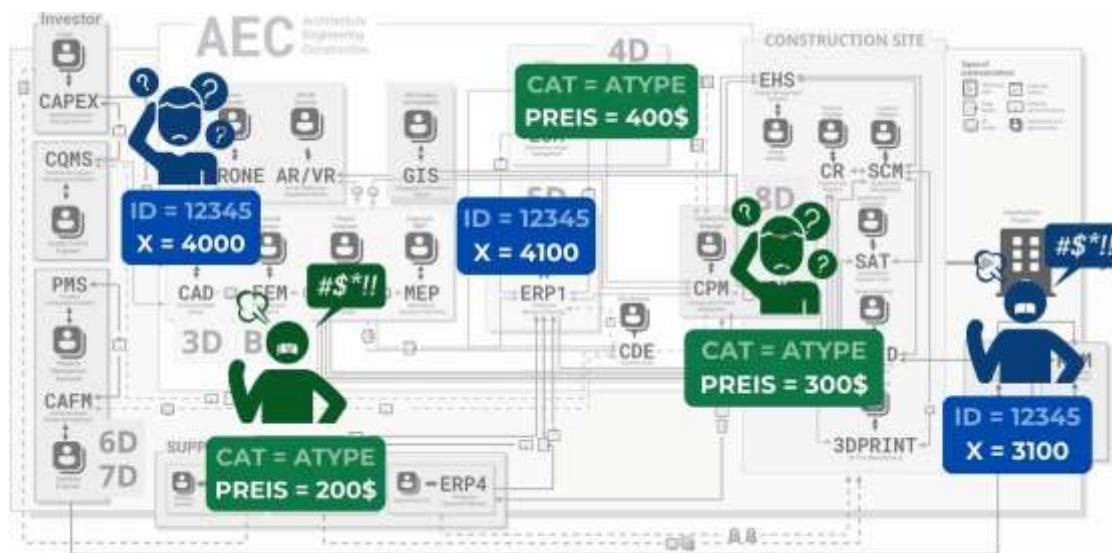


Fig. 2.1-6 I arbeidet med å finne de riktige dataene må ledere sikre kvaliteten og den juridiske påliteligheten til data mellom ulike systemer.

I praksis skaper imidlertid denne modellen nye problemer. Når data håndteres manuelt og informasjonen er spredt over mange beslutninger som ikke har noe med hverandre å gjøre, blir ethvert forsøk på å skaffe nøyaktig og oppdatert informasjon gjennom en pyramide av beslutningstakere (figur 2.1-7) en flaskehals - tidkrevende og feilutsatt.

Situasjonen forverres av et skred av digitale løsninger. Programvaremarkedet oversvømmes stadig av nye verktøy som virker lovende. Men uten en klar strategi for datahåndtering integreres ikke disse løsningene i et enhetlig system, men skaper i stedet flere lag med kompleksitet og duplisering. I stedet for å forenkle prosessene, ender bedriftene derfor opp i et enda mer fragmentert og kaotisk informasjonsmiljø.

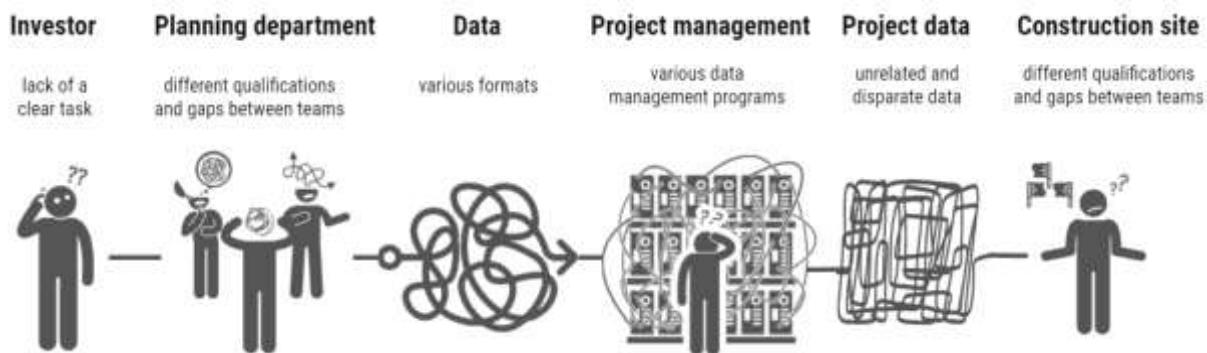


Fig. 2.1-7 Kompleksiteten i systemene og de mange ulike dataformatene fører til at konsistensen i byggeprosessen går tapt.

Alle disse problemene knyttet til håndtering av et mangfold av ulike løsninger før eller siden til en viktig erkjennelse hos bedriftsledelsen: Det handler ikke om datamengden eller om å finne det neste "one-size-fits-all"-verktøyet for å behandle den. Den virkelige årsaken ligger i kvaliteten på dataene og hvordan

organisasjonen skaper, mottar, lagrer og bruker dem.

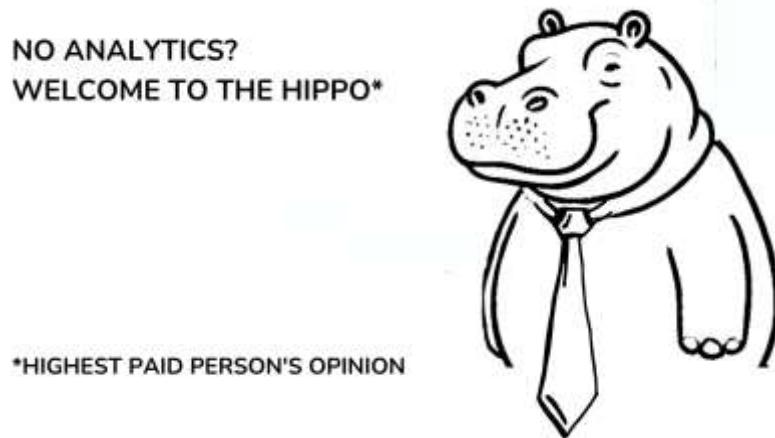
Nøkkelen til bærekraftig suksess ligger ikke i å jakte på nye "magiske" applikasjoner, men i å bygge en datakultur i selskapet. Det innebærer å behandle data som en strategisk ressurs og prioritere datakvalitet, -integritet og -relevans på alle nivåer i organisasjonen.

Løsningen på dilemmaet mellom kvalitet og kvantitet ligger i å skape en enhetlig datastruktur som eliminerer duplisering, eliminerer inkonsekvenser og forenkler informasjonsflyten. Denne arkitekturen gir én enkelt, pålitelig datakilde som kan brukes til å ta informerte, nøyaktige og tidsriktige beslutninger.

Ellers fortsetter selskapene å stole på HiPPO-ekspertenes subjektive meninger og intuitive vurderinger i stedet for på pålitelige fakta, noe som fortsatt ofte er tilfelle. I byggebransjen, der ekspertise tradisjonelt spiller en viktig rolle, er dette spesielt merkbart.

HiPPO eller faren for meninger i beslutningsprosesser

Tradisjonelt sett tas viktige beslutninger i byggebransjen basert på erfaring og subjektive vurderinger. Uten pålitelige data i rett tid må bedriftslederne handle i blinde og stole på intuisjonen til de best betalte medarbeiterne (HiPPO - Highest Paid Person's Opinion) i stedet for på objektive fakta (fig. 2.1-8).



Figur 2.1-8 I mangel av analyseverktøy er virksomheten avhengig av erfarne fagfolks subjektive vurderinger.

Denne tilnærmingen kan være berettiget i et stabilt og langsomt skiftende miljø, men i en tid preget av digital transformasjon utgjør den en alvorlig risiko. Beslutninger basert på intuisjon og gjetning er utsatt for forvrengninger, er ofte basert på hypoteser som ikke er underbygget, og tar ikke hensyn til det komplekse bildet som gjenspeiles i dataene

Det som utgis for å være en intelligent debatt på beslutningsnivå i en bedrift, er ofte ikke basert på noe konkret. En bedrifts suksess bør ikke være avhengig av eksperters autoritet og lønn, men av evnen til å arbeide effektivt med data, identifisere mønstre og ta informerte beslutninger.

Det er viktig å forlate forestillingen om at autoritet eller erfaring automatisk betyr at en beslutning er riktig. Den datadrevne tilnærmingen er en game changer: Data og analyser, ikke stilling og lønn, er nå grunnlaget for beslutningstaking. Stordata, maskinlæring og visuell analyse gjør det mulig for oss å identifisere mønstre og basere oss på fakta i stedet for gjetninger (fig. 1.1-4).

Uten data er du bare enda en person med en mening [34].

- W. Edwards Deming, forsker og ledelseskonsulent

Moderne datahåndteringsmetoder sikrer også kontinuitet i kunnskapen innad i bedriften. Tydelig beskrevne prosesser, automatisering og en systematisk tilnærming gjør det mulig å overføre selv nøkkelroller uten å miste effektivitet.

Men blind tillit til data kan også føre til alvorlige feil. Data i seg selv er bare en samling tall. Uten riktig analyse, kontekst og evnen til å identifisere mønstre har de ingen verdi og kan ikke drive prosesser. Nøkkelen til suksess ligger ikke i å velge mellom HiPPO-intuisjon og analyse, men i å bygge intelligente verktøy som forvandler uensartet informasjon til håndterbare, informerte beslutninger.

I digitalt byggemiljø er det ikke ansiennitet og plassering i hierarkiet som blir avgjørende suksessfaktorer, men reaksjonsevne, beslutningsnøyaktighet og ressurseffektivitet.

Data er verktøy, ikke absolutte sannheter. De bør utfylle menneskelig tenkning, ikke erstatte den. Til tross for fordelene med analyser, kan ikke data erstatte menneskelig intuisjon og erfaring. Deres rolle er å bidra til å ta mer nøyaktige og informerte beslutninger.

Konkurransefortrinn oppnås ikke bare ved å oppfylle standarder, men også ved å utkonkurrere konkurrentene i effektiv bruk av ressurser som er de samme for alle. I fremtiden vil datakompetanse bli like viktig som lese-, skrive- og regneferdigheter en gang var. Fagfolk som kan analysere og tolke data, vil kunne ta mer nøyaktige beslutninger og erstatte dem som kun baserer seg på personlig erfaring (figur 2.1-9).

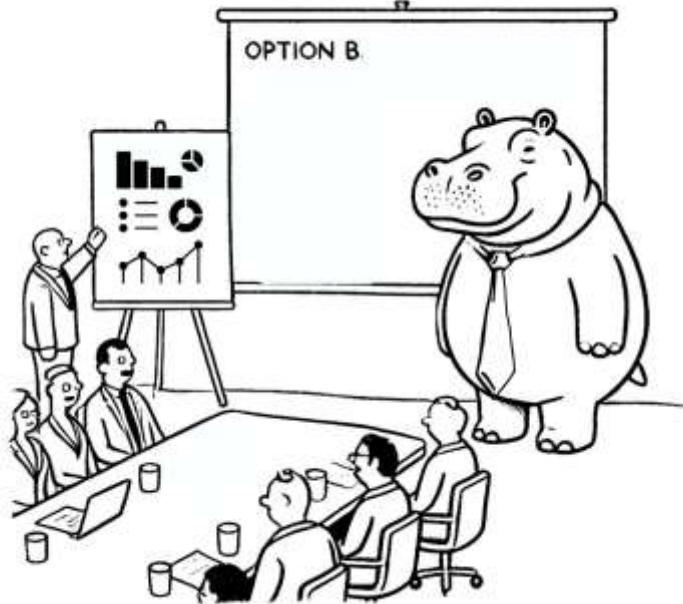


Fig. 2.1-9 Beslutninger bør være basert på objektive analyser, ikke på hva den best betalte medarbeideren mener.

Ledere, spesialister og ingeniører vil fungere som dataanalytikere på og studere prosjektenes struktur, dynamikk og nøkkelindikatorer. De menneskelige ressursene vil bli en del av systemet, noe som krever fleksibel, datadrevet tilpasning for å maksimere effektiviteten.

Feilene ved bruk av utilstrekkelige data er mye mindre enn når man ikke bruker data [35].

- Charles Babbage, oppfinneren av den første analytiske regnemaskinen

Fremveksten av stordata og innføringen av LLM (Large Language Models) har radikalt endret ikke bare måten vi analyserer på, men også selve beslutningsprosessen. Mens man tidligere fokuserte på årsakssammenhenger (hvorfor noe skjedde - diagnostisk analyse) (fig. 1.1-4), er det i dag evnen til å forutsi fremtiden (prediktiv analyse) og, i fremtiden, preskriptiv analyse, der maskinlæring og kunstig intelligens foreslår det beste valget i beslutningsprosessen, som kommer i forgrunnen.

Ifølge den nye SAP™-studien "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36], ville 44% av topplederne være villige til å endre sin tidligere beslutning basert på råd fra kunstig intelligens, og 38% ville stole på at kunstig intelligens tok forretningsbeslutninger på deres vegne. Samtidig sier 74% av lederne at de stoler mer på råd fra kunstig intelligens enn på råd fra venner og familie, og 55% jobber i selskaper der kunstig intelligens erstatter eller ofte går utenom tradisjonelle beslutningsmetoder - spesielt i organisasjoner med en årlig omsetning på over 5 milliarder dollar. I tillegg bruker 48% av respondentene generative AI-verktøy daglig, og 15% av dem bruker dem flere ganger om dagen.

Med utviklingen av LLM og automatiserte datahåndteringssystemer oppstår det en ny utfordring: hvordan bruke informasjon effektivt uten å miste verdien i kaoset av inkompatible formater og heterogene kilder, noe som suppleres av den økende kompleksiteten og dynamikken i forretningsprosessene.

Kontinuerlig økning i kompleksiteten og dynamikken i forretningsprosessene

Byggebransjen står i dag overfor alvorlige utfordringer når det gjelder data- og prosesshåndtering. Hovedutfordringene er fragmenterte informasjonssystemer, overdrevent byråkrati og manglende integrasjon mellom digitale verktøy. Disse utfordringene forsterkes i takt med at forretningsprosessene i seg selv blir mer komplekse - drevet av teknologi, endrede kundekrav og regelverk i stadig utvikling.

At byggeprosjekter er unike, skyldes ikke bare deres tekniske særegheter, men også forskjeller i nasjonale standarder og myndighetskrav i ulike land (fig. 4.2-10, fig. 5.1-7). Dette krever en fleksibel, individualisert tilnærming til hvert enkelt prosjekt, noe som er vanskelig å gjennomføre med tradisjonelle modulære kontrollsystemer. På grunn av kompleksiteten i prosessene og de store datamengdene henvender mange bedrifter seg til leverandører som tilbyr spesialiserte løsninger. Men markedet er overbelastet - mange nystartede selskaper tilbyr lignende produkter med fokus på smale oppgaver. Resultatet er at man ofte mister den helhetlige tilnærmingen til datahåndtering.

Tilpasning til den kontinuerlige strømmen av ny teknologi og nye markedskrav er i ferd med å bli en kritisk faktor for konkurransen. Eksisterende proprietære applikasjoner og modulære systemer har imidlertid lav tilpasningsevne - eventuelle endringer krever ofte langvarige og kostbare revisjoner av utviklere som ikke alltid forstår detaljene i byggeprosessene.

Bedriftene blir tatt som gisler av teknologisk etterslep, og venter på nye oppdateringer i stedet for å implementere innovative, integrerte tilnæringer raskt. Resultatet er at den interne strukturen i bygge- og anleggsorganisasjoner ofte er et komplekst økosystem av sammenkoblede hierarkiske, og ofte lukkede, systemer som koordineres gjennom et nettverk av ledere på flere nivåer (figur 2.1-10).

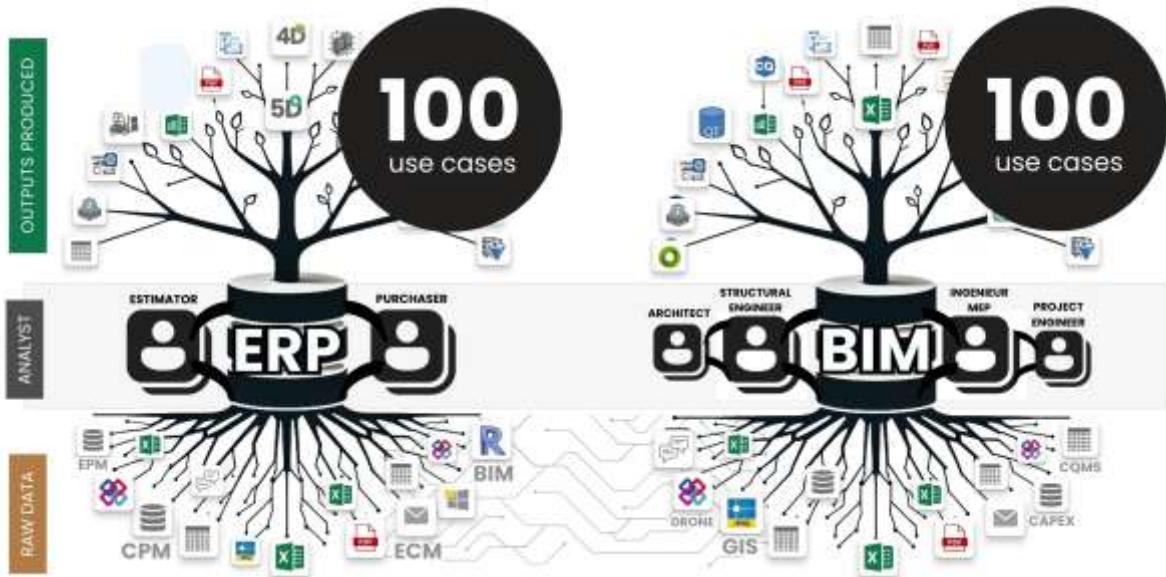


Fig. 2.1-10 Bedrifter består av sammenkoblede systemer som sammen danner prosesser som krever automatisering.

Ifølge en studie utført av Canadian Construction Association og KPMG Canada 2021 [37], mener bare 25% av selskapene at de er i en betydelig eller annerledes posisjon sammenlignet med konkurrentene når det gjelder bruk av teknologi eller digitale løsninger. Bare 23% av respondentene rapporterte at løsningene deres i stor eller svært stor grad er datadrevne. Samtidig karakteriserte flertallet av respondentene i undersøkelsen bruken av en rekke andre teknologier som rent eksperimentell, eller innrømmet at de ikke bruker dem i det hele tatt.

Denne motviljen mot å delta i teknologiske eksperimenter er spesielt tydelig i store infrastrukturprosjekter, der feil kan koste millioner av dollar. Selv de mest avanserte teknologiene - digitale tvillinger, prediktiv analyse - møter ofte motstand, ikke på grunn av effektiviteten, men fordi de ikke har vist seg å være pålitelige i reelle prosjekter.

Ifølge rapporten "Shaping the Future of Construction" [5] fra World Economic Forum (WEF) står innføringen av ny teknologi i byggebransjen ikke bare overfor tekniske vanskeligheter, men også psykologiske barrierer fra kundenes side [5] står innføringen av ny teknologi i byggebransjen ikke bare overfor tekniske vanskeligheter, men også psykologiske barrierer fra kundenes side. Mange kunder frykter at bruken av avanserte løsninger vil gjøre prosjektene deres til et eksperimentområde og gjøre dem til "forsøkskaniner", og uforutsigbare konsekvenser kan føre til ekstra kostnader og risiko.

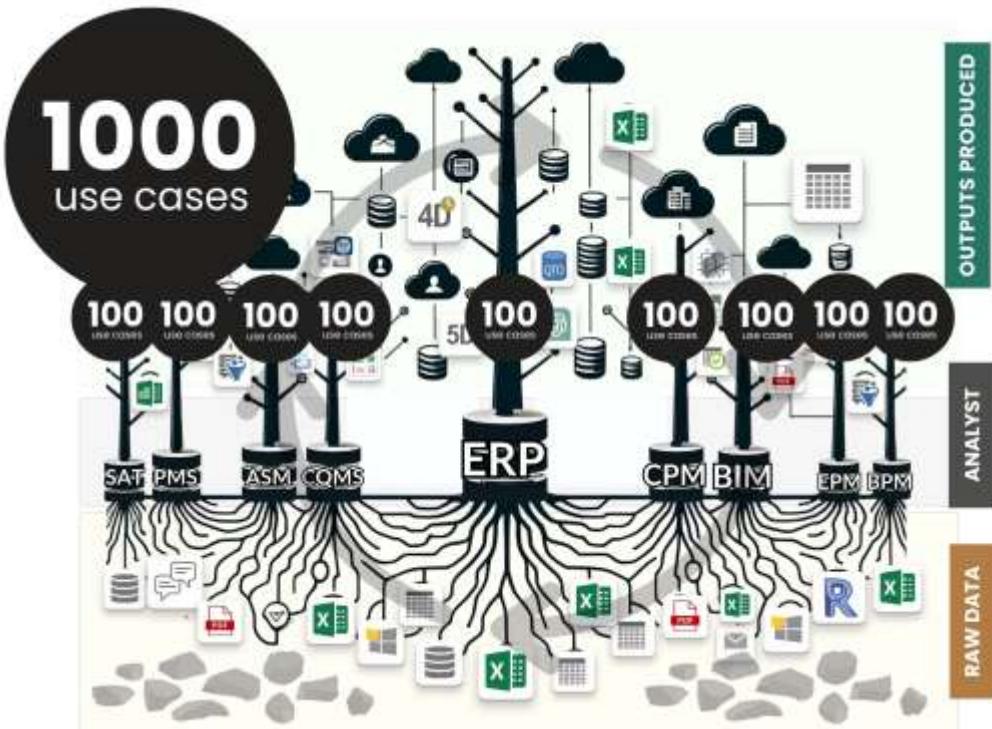


Fig. 2.1-11 For hvert dataanvendelsestilfelle tilbyr løsningsmarkedet applikasjoner for optimalisering og automatisering av prosesser.

Byggebransjen er svært mangfoldig: ulike prosjekter har ulike krav, regionale særegenheter, lovbestemte klassifiseringsforskrifter (fig. 4.2-10), beregningsstandarder (fig. 5.1-7) osv. Derfor er det praktisk talt umulig å lage en proprietær universell applikasjon eller et system som passer perfekt til alle disse kravene og prosjektspesifikasjonene.

I et forsøk på å håndtere den økende kompleksiteten i systemene og avhengigheten av programvareleverandører, innser man i økende grad at nøkkelen til effektiv datahåndtering ikke bare er åpenhet og standardisering, men også forenkling av selve prosessarkitekturen. Den økende kompleksiteten og dynamikken i forretningsprosessene krever nye tilnærmingar, der prioriteten skifter fra å samle data til å strukturere og organisere dem. Det er dette skiftet som vil være det neste steget i utviklingen av byggebransjen, og som vil markere slutten på tiden med programvareleverandørenes dominans og begynnelsen på en tid med meningsfull informasjonsorganisering.

Erkjennelsen av begrensningene ved universalløsninger og sårbarheten overfor økende kompleksitet fører til et skifte i prioriteringene fra lukkede plattformer og datahøsting til åpenhet, tilpasningsdyktighet og struktureret informasjonshåndtering. Dette skiftet i tenkemåte gjenspeiler bredere endringer i den globale økonomien og teknologien, som beskrives gjennom såkalte "industrielle revolusjoner". For å forstå hvor bygg- og anleggsbransjen er på vei og hvilken retning den vil ta i fremtiden, er det nødvendig å vurdere bransjens plass i den fjerde og femte industrielle revolusjonen - fra automatisering og digitalisering til personalisering, åpne standarder og den tjenestebaserte datamodellen.

Den fjerde industrielle revolusjonen (Industri 4.0) og den femte industrielle revolusjonen (Industri 5.0) i bygg- og anleggsbransjen

Teknologiske og økonomiske stadier er teoretiske begreper som brukes til å beskrive og analysere samfunnets og økonomiens utvikling på ulike utviklingstrinn. De kan tolkes ulikt av ulike forskere og eksperter.

- **Den fjerde industrielle revolusjon** (4IR eller Industry 4.0) er knyttet til informasjonsteknologi, automatisering, digitalisering og globalisering. Et av de viktigste elementene er utviklingen av proprietære programvareløsninger, det vil si spesialiserte digitale produkter som er utviklet for spesifikke oppgaver og bedrifter. Disse løsningene blir ofte en viktig del av IT-infrastrukturen, men er lite skalerbare uten ytterligere modifikasjoner.
- **Den femte industrielle revolusjonen (5IR)** er nå på et tidligere stadium av konseptualisering og utvikling enn 4IR. Kjerneprinsippene er blant annet økt personalisering av produkter og tjenester. 5IR er en bevegelse mot mer tilpasningsdyktig, fleksibel og persontilpasset økonomisk aktivitet med fokus på persontilpasning, rådgivning og tjenesteorienterte modeller. Et sentralt aspekt ved den femte økonomiske modusen er bruk av data i beslutningsprosesser, noe som er praktisk talt umulig uten bruk av åpne data og åpne verktøy (figur 2.1-12).

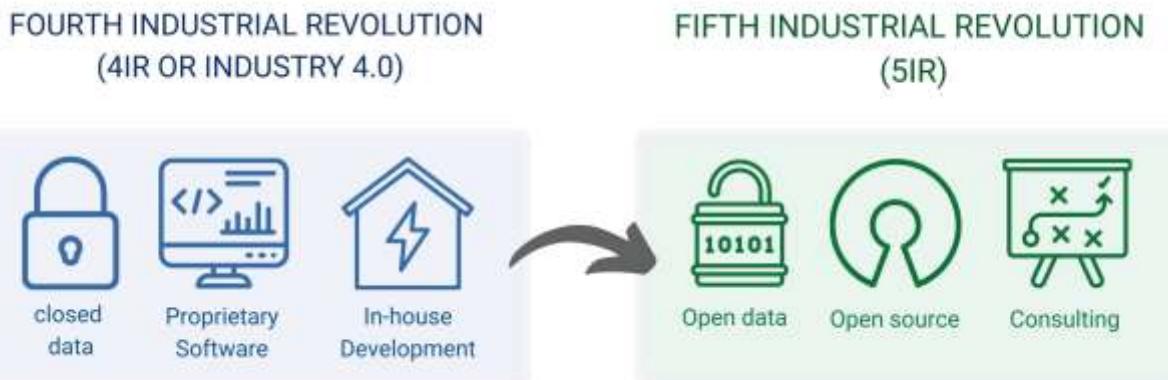


Fig. 2.1-12 Det fjerde mønsteret fokuserer på løsninger, mens det femte mønsteret fokuserer på personalisering og data.

Å lage en applikasjon for bedrifter i byggebransjen som kan brukes i ti eller hundre organisasjoner, er ingen garanti for at den kan skaleres til andre bedrifter, regioner eller land uten betydelige endringer og forbedringer. Sannsynligheten for å lykkes med å skalere slike løsninger er fortsatt lav, ettersom hver organisasjon har unike prosesser, krav og forhold som kan kreve personlige tilpasninger.

Det er viktig å forstå at en vellykket integrering av teknologiske løsninger allerede i dag forutsetter en dypt personlig tilnærming til hver enkelt prosess, prosjekt og bedrift. Det betyr at selv etter at et universelt rammeverk, verktøy eller program er utviklet, må det tilpasses og skreddersys i detalj for å oppfylle de unike kravene og betingelsene som gjelder for hver enkelt bedrift og hvert enkelt prosjekt.

Ifølge PwC-rapporten "Decoding the Fifth Industrial Revolution" [38] er rundt 50% av topplederne i ulike bransjer i år avhengige av å integrere avansert teknologi og menneskelig ekspertise. Denne tilnærmingen gjør at de raskt kan tilpasse seg endringer i produktdesign eller kundekrav og skape persontilpasset produksjon.

Hver prosess krever utvikling av en unik funksjon eller applikasjon, noe som, med tanke på størrelsen på den globale byggebransjen og mangfoldet av prosjekter, fører til at det finnes et stort antall business cases, som hver gang representerer en unik Pipeline logikk (fig. 2.1-13). Hvert av disse casene har sine egne særegenheter og krever en skreddersydd tilnærming. Vi vil se nærmere på de mange mulige løsningene på det samme analytiske problemet i sammenheng med ulike tilnærminger i kapittelet om maskinlæring og analysering av Titanic-datasettet (fig. 9.2-9).

Pipeline i forbindelse med digitale prosesser er en sekvens av aktiviteter, prosesser og verktøy som muliggjør en automatisert eller strukturert flyt av data og arbeid gjennom de ulike fasene i prosjektets livssyklus.

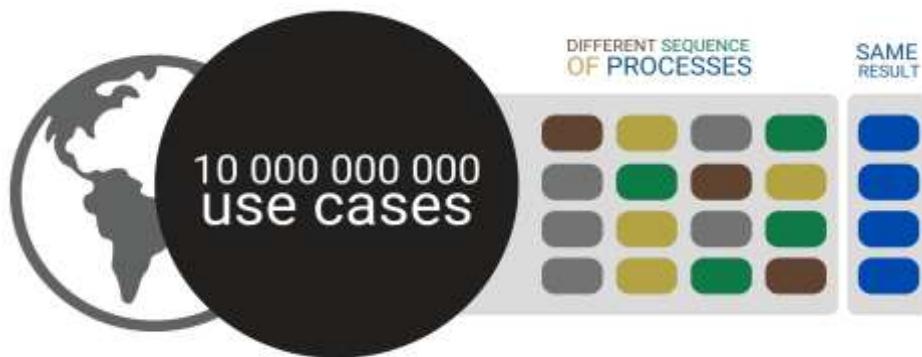


Fig. 2.1-13 Forretningssakenes individualitet og variasjon gjør det umulig å skape skalerbare, lukkede plattformer og verktøy.

Våre liv har allerede endret seg på mange måter under påvirkning av den digitale transformasjonen, og i dag kan vi snakke om en ny fase i den økonomiske utviklingen av byggebransjen. I denne "nye økonomien" vil konkurransen organiseres etter andre regler: Den som er i stand til å omdanne offentlig kunnskap og åpne data til etterspurte produkter og tjenester på en effektiv måte, får en viktig fordel i den femte industrielle revolusjonen.

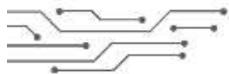
Som økonomen Kate Maskus bemerker i boken "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39] 2012 [39], "lever vi i en global kunnskapsøkonomi, og fremtiden tilhører dem som vet hvordan man gjør vitenskapelige oppdagelser om til handelsvarer".

Overgangen til den femte økonomiske modusen innebærer et skifte i fokus fra lukkede IT-løsninger til åpne standarder og plattformer. Selskapene vil begynne å gå bort fra tradisjonelle programvareprodukter til fordel for tjenesteorienterte modeller, der hovedaktiviteten vil være data i stedet for proprietær teknologi.

Harvard Business School 2024-studien [40] viser den enorme økonomiske verdien av programvare med åpen kildekode (Open Source Software, OSS). Ifølge studien finnes OSS i 96% av all programvarekode, og enkelte kommersielle programvarer består av 99,9% OSS-komponenter. Uten OSS ville bedriftene brukt 3,5 ganger så mye penger på programvare.

I tråd med de globale trendene vil oppbyggingen av bedriftsøkosystemer gradvis gå over i et femte økonomisk paradigme, der datasentriske analyse- og konsulenttjenester vil få høyere prioritet enn isolerte, lukkede løsninger med strengt definerte bruksscenarioer.

Digitaliseringens tidsalder vil endre maktbalansen i bransjen: I stedet for å basere seg på leverandør løsninger vil selskapene basere sin konkurranseevne på sin evne til å bruke data effektivt. Som et resultat av dette vil byggebransjen gå fra gamle, rigide systemer til fleksible, tilpasningsdyktige økosystemer der åpne standarder og interoperable verktøy vil danne grunnlaget for prosjektstyringen. Slutten på en æra med applikasjonsleverandørdominans vil skape et nytt miljø der verdi ikke defineres ut fra besittelse av lukket kildekode og spesialiserte konnektorer, men ut fra evnen til å gjøre data til en strategisk fordel.



KAPITTEL 2.2.

GJØRE KAOS TIL ORDEN OG REDUSERE KOMPLEKSITETEN

Overflødig kode og lukkede systemer hindrer produktivitetsforbedringer

De siste tiårene har de teknologiske endringene på IT-området først og fremst vært drevet av programvareleverandørene. De har satt kurset for utviklingen og bestemt hvilke teknologier bedriftene skulle ta i bruk, og hvilke de skulle legge bak seg. I en tid da man gikk fra siloløsninger til sentraliserte databaser og integrerte systemer, markedsførte leverandørene lisensierte produkter som ga kontroll over tilgang og skalerbarhet. Senere, med inntoget av skyteknologi og SaaS-modeller (Software as a Service), utviklet denne kontrollen seg til en abonnementsmodell som gjorde brukerne til lojale kunder av digitale tjenester.

Denne tilnærmingen har ført til et paradoks: Til tross for de enorme mengdene programkode som skapes, er det bare en liten del av den som faktisk blir brukt. Kanskje er det hundrevis eller tusenvis av ganger mer kode enn nødvendig, fordi de samme forretningsprosessene er beskrevet og duplisert i dusinvise eller hundrevis av programmer på forskjellige måter, til og med innenfor samme selskap. Samtidig er utviklingskostnadene allerede betalt, og disse kostnadene er u gevinnelige. Likevel fortsetter bransjen å reproduksjonse denne syklusen, og skaper nye produkter med minimal merverdi for sluttbrukeren, oftere under press fra markedets forventninger enn reelle behov.

Ifølge Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41] kan kostnadene for programvareutvikling variere betydelig avhengig av flere faktorer, blant annet systemets kompleksitet og hvilken teknologi som er valgt. Historisk sett har utviklingskostnadene i 2008 ligget på rundt 100 dollar per kildekodelinje (SLOC), mens vedlikeholdskostnadene kan stige til 4 000 dollar per SLOC.

Bare én av komponentene i CAD-applikasjoner - den geometriske kjernen - kan bestå av flere titalls millioner kodelinjer (fig. 6.1-5). En lignende situasjon finner vi i ERP-systemer (fig. 5.4-4), som vi kommer tilbake til i bokens femte del for å diskutere kompleksiteten i. En nærmere titt avslører imidlertid at mye av denne koden ikke tilfører verdi, men bare fungerer som et "postbud" - som mekanisk flytter data mellom databasen, API-et, brukergrensesnittet og andre tabeller i systemet. Til tross for den populære myten om den kritiske betydningen av såkalt forretningslogikk, er den harde virkeligheten mye mer prosaisk: Moderne kodebasar er fulle av utdaterte malblokker (legacy-kode) hvis eneste formål er å sikre dataoverføring mellom tabeller og komponenter uten å påvirke beslutningstaking eller forretningseffektivitet.

Resultatet er at lukkede løsninger som behandler data fra ulike kilder, uunngåelig blir til uoversiktlige "spaghetti-økosystemer". Disse komplekse, sammenflettede systemene kan bare håndteres av en hærskare av ledere som jobber semi-rutinemessig. Denne måten å organisere datahåndteringen på er ikke bare ressursløsende, men skaper også kritiske sårbarheter i forretningsprosessene, noe som gjør bedriften avhengig av en snever krets av spesialister som forstår hvordan denne teknologiske labyrinten fungerer.

Den kontinuerlige økningen i mengden kode, antall applikasjoner og den økende kompleksiteten i konseptene som tilbys av leverandørene, har ført til et naturlig resultat - en økning i kompleksiteten i IT-økosystemet i bygg- og anleggsbransjen. Dette har gjort den praktiske gjennomføringen av

digitaliseringen ved å øke antallet applikasjoner i bransjen ineffektiv. Programvareprodukter som er laget uten å ta hensyn til brukernes behov, krever ofte betydelige ressurser til implementering og support, men gir ikke den forventede avkastningen.

Ifølge McKinseys studie "Increasing Construction Productivity" [42] har den globale veksten i arbeidsproduktiviteten i bygg- og anleggsbransjen i løpet av de siste to tiårene i gjennomsnitt bare vært på 1% per år, sammenlignet med en vekst på 2,8% for verdensøkonomien som helhet og 3,6% for industrien. I USA har arbeidsproduktiviteten per arbeider i bygg- og anleggsbransjen blitt halvert siden 1960-tallet [43].

Stadig mer komplekse systemer, isolasjon og lukkede data har svekket kommunikasjonen mellom fagfolk, noe som gjør byggebransjen til en av de minst effektive (figur 2.2-1). til 22 000 milliarder dollar innen 2040, noe som vil kreve betydelige effektiviseringsgevinster.

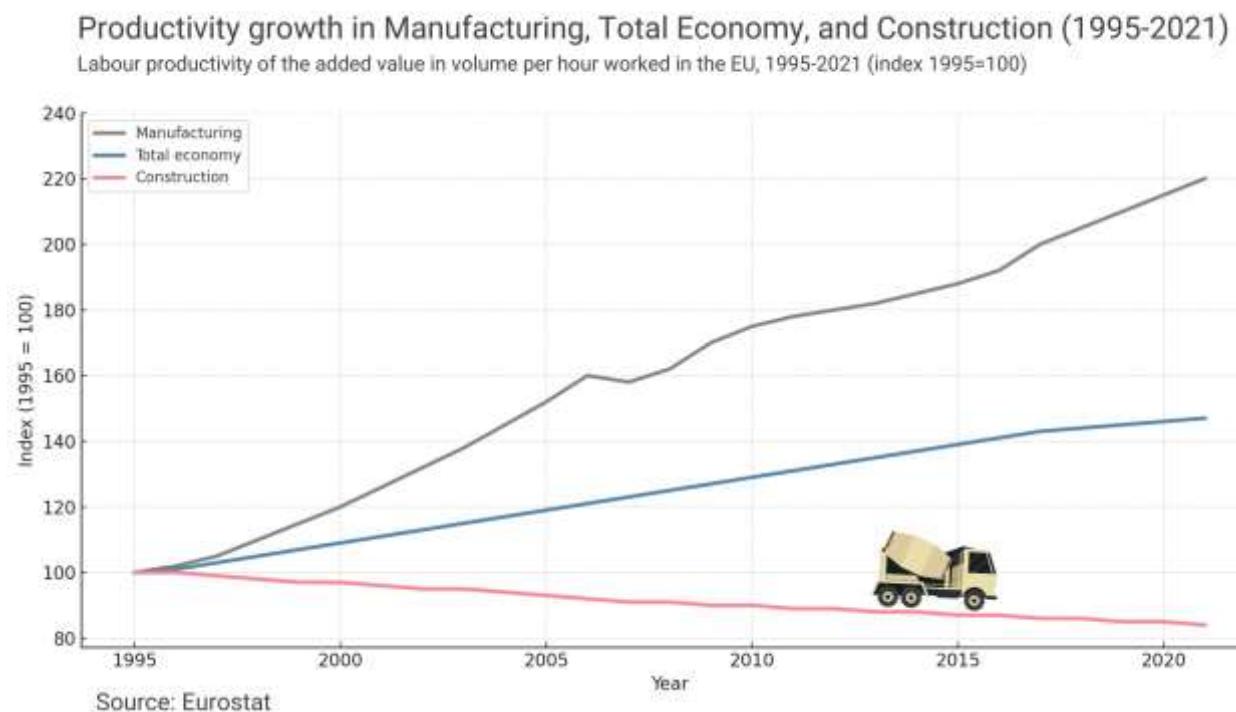


Fig. 2.2-1 Lukkede og komplekse data og følgelig dårlig kommunikasjon mellom spesialister har gjort byggebransjen til en av de minst effektive sektorene i økonomien (basert på [44], [45]).

Som det understres i McKinsey-studien "Ensuring construction productivity is no longer optional" (2024), har byggebransjen ikke lenger råd til å holde seg på dagens produktivitetsnivå på grunn av den økende ressursknappheten og bransjens ønske om å doble vekstraten [44]. De globale byggekostnadene forventes å stige fra 13 000 milliarder dollar i 2023 til mye høyere nivåer innen utgangen av tiåret, noe som gjør spørsmålet om effektivitet ikke bare relevant, men kritisk.

En av de viktigste måtene å forbedre effektiviteten på vil være den unngåelige foreningen og forenklingen av applikasjonsstrukturer og dataøkosystemarkitekturen. Denne tilnærmingen til rasjonalisering vil eliminere overflødige abstraksjonslag og unødvendig kompleksitet som har høpt seg opp gjennom årene i bedriftens systemer.

Fra siloer til ett enkelt datavarehus

Jo mer data en organisasjon samler på seg, desto vanskeligere blir det å hente ut reell verdi fra dem. På grunn av den fragmenterte lagringen av informasjon i isolerte siloer er moderne bedrifters forretningsprosesser som byggmestere som prøver å bygge en skyskaper av materialer som er lagret i tusenvis av forskjellige lagerbygninger. Overskuddet av informasjon gjør det ikke bare vanskelig å få tilgang til juridisk relevant informasjon, men bremser også beslutningsprosessen: Hvert trinn må kontrolleres og bekreftes gjentatte ganger.

Hver oppgave eller prosess er koblet til en egen tabell eller database, og datautveksling mellom systemer krever komplekse integrasjoner. Feil og inkonsekvenser i ett system kan føre til kjedefeil i andre. Feilaktige verdier, sene oppdateringer og duplisert informasjon tvinger de ansatte til å bruke mye tid på manuell avstemming og sammenstilling av data. Resultatet er at organisasjonen bruker mer tid på å håndtere konsekvensene av fragmenteringen enn å utvikle og optimalisere prosessene.

Dette problemet er universelt: Noen bedrifter fortsetter å slite med kaos, mens andre finner en løsning i integrasjon - å flytte informasjonsflyten inn i et sentralisert lagringssystem. Tenk på det som én stor tabell der du kan lagre alle enheter knyttet til oppgaver, prosjekter og objekter. I stedet for dusinvis av ulike tabeller og formater oppstår det ett sammenhengende arkiv (figur 2.2-2) som gjør det mulig:

- minimere tap av data;
- eliminere behovet for konstant harmonisering av informasjon;
- forbedre datatilgjengeligheten og -kvaliteten;
- forenkle analytisk behandling og maskinlæring

Ved å bringe data til en felles standard blir informasjonen konvertert til et enhetlig og maskinlesbart format, uavhengig av kilde. En slik organisering av data gjør det mulig å kontrollere integriteten, analysere dem i sanntid og bruke dem til å ta ledelsesbeslutninger.

Konseptet med integrerte lagringssystemer og deres anvendelse innen analyse og maskinlæring vil bli diskutert mer inngående i kapittelet "Big Data Storage and Machine Learning". Temaene datamodellering og -strukturering dekkes i detalj i kapitlene "Transforming data into a structured form" og "How standards

change the game: from random files to an elaborate data model".

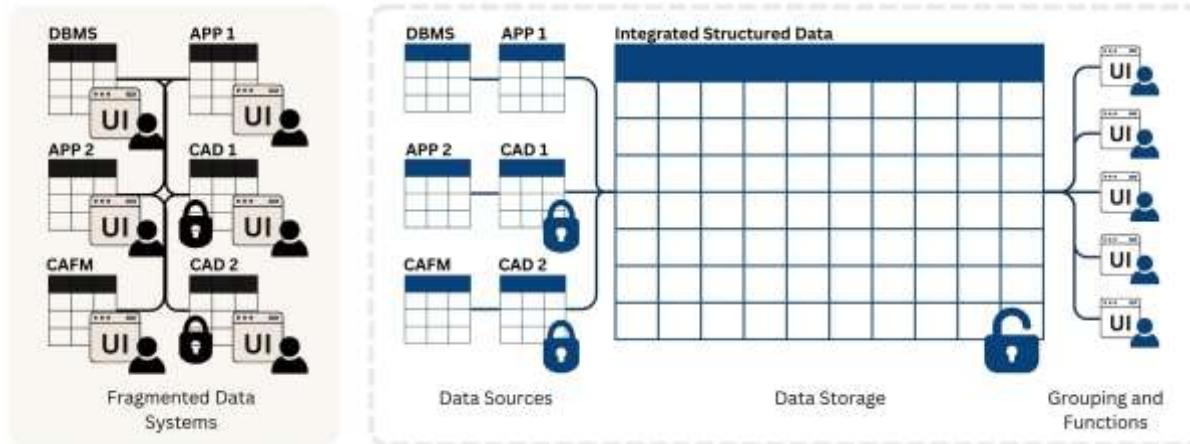


Fig. 2.2-2 Dataintegrasjon eliminerer siloer, forbedrer informasjonstilgjengeligheten og optimaliserer forretningsprosessene.

Når dataene er strukturert og sammenslått, er det neste logiske trinnet å validere dem. Med ett enkelt integrert datalager blir denne prosessen svært forenklet: Det er slutt på flere inkonsekvente skjemaer, dupliserte strukturer og komplekse relasjoner mellom tabeller. All informasjon er tilpasset en enkelt datamodell, noe som eliminerer interne inkonsistenser og gjør valideringsprosessen raskere. Validering og sikring av datakvalitet er hjørnestensaspekter i alle forretningsprosesser, og vi vil se nærmere på dem i de relevante kapitlene i boken.

I siste fase grupperes, filtreres og analyseres dataene. De blir behandlet med ulike funksjoner: aggregering (addisjon, multiplikasjon), beregninger mellom tabeller, kolonner eller rader (fig. 2.2-4). Arbeidet med data blir en sekvens av trinn: innsamling, strukturering, validering, transformasjon, analytisk behandling og overføring til endelige applikasjoner der informasjonen brukes til å løse praktiske problemer. Vi kommer til å diskutere mer om hvordan man bygger slike scenarier, automatiserer trinn og bygger prosesseringsflyter i kapitlene om ETL -prosesser og datapipeline-tilnærming.

Digital transformasjon handler derfor ikke bare om å forenkle håndteringen av informasjon. Det handler om å eliminere overdreven kompleksitet i datahåndteringen, å gå fra kaos til forutsigbarhet, fra flere systemer til en håndterbar prosess. Jo mindre kompleks arkitekturen er, desto mindre kode kreves det for å støtte den. Og i fremtiden kan koden som sådan forsvinne helt og gi plass til intelligente agenter som på egen hånd analyserer, systematiserer og transformerer data.

Integrerte lagringssystemer muliggjør overgangen til AI agenter

Jo mindre komplekse dataene og systemene er, desto mindre kode trenger du å skrive og vedlikeholde. Og den enkleste måten å spare utvikling på er å kvitte seg helt med kode og erstatte den med data. Når applikasjonsutvikling går fra kode til datamodeller, skjer det unngåelig et skifte mot en datasentrisk (datadrevet) tilnærming, fordi det er en helt annen måte å tenke på som ligger bak disse konseptene.

Når man velger å jobbe med data i sentrum, begynner man å se annerledes på deres rolle. Data er ikke lenger bare "råmateriale" for applikasjoner - det er nå fundamentet som arkitektur, logikk og interaksjon bygges rundt.

Den tradisjonelle tilnærmingen til datahåndtering starter vanligvis på applikasjonsnivå og ligner på et tungvint byråkratisk system: godkjenninger på flere nivåer, manuelle kontroller, endeløse versjoner av dokumenter gjennom relevante programvareprodukter. Med utviklingen av digital teknologi vil flere og flere selskaper bli tvunget til å gå over til prinsippet om minimalisme - å lagre og bruke bare det som virkelig er nødvendig og som vil bli brukt.

Logikken om minimering har blitt tatt opp av leverandørene. For å forenkle datalagring og -behandling flyttes brukernes arbeid fra offline-applikasjoner og -verktøy til skytjenester og såkalte SaaS-løsninger.

SaaS-konseptet (Software as a Service, eller "programvare som en tjeneste") er en av de viktigste trendene innen moderne IT-infrastrukturer, og gjør det mulig for brukere å få tilgang til applikasjoner via Internett uten å måtte installere og vedlikeholde programvare på sine egne datamaskiner.

På den ene siden har SaaS gjort det enklere å skalere, versjonskontrollere og redusere support- og vedlikeholdskostnader, men på den andre siden har det også gjort brukeren helt avhengig av leverandørens skyinfrastruktur, i tillegg til å være avhengig av logikken i en bestemt applikasjon. Hvis en tjeneste går ned, kan tilgangen til data og forretningsprosesser bli midlertidig eller til og med permanent blokkert. I tillegg lagres alle brukerdata i SaaS-applikasjoner på leverandørens servere, noe som utgjør en risiko for sikkerhet og overholdelse av regelverk. Endringer i tariffer eller bruksvilkår kan også føre til økte kostnader eller behov for hurtig migrering.

Utviklingen av AI, LLM -agenter og datasentrisk tilnærming har satt spørsmålstege ved fremtiden for applikasjoner i deres tradisjonelle form og SaaS -utførelse. Mens applikasjoner og tjenester tidligere måtte håndtere forretningslogikk og behandle data, kan disse funksjonene med AI-agenter flyttes over til intelligente systemer som arbeider direkte med data.

Derfor diskutes det i økende grad hybridarkitekturen i IT-avdelingene og på ledelsesnivå, der AI -agenter og lokale løsninger supplerer skytjenester og reduserer avhengigheten av SaaS -plattformer.

Vår tilnærming tar høyde for at tradisjonelle forretningsapplikasjoner eller SaaS-applikasjoner -applikasjoner kan endre seg dramatisk i agentenes tidsalder. Disse applikasjonene er i hovedsak CRUD [create, read, update and delete]-databaser med forretningslogikk. Men i fremtiden vil denne logikken bli overtatt av AI-agenter [46].

- Satya Nadella, administrerende direktør i Microsoft, 2024.

En datasentrisk tilnærming og bruk av AI/LLM-agenter kan redusere overflødige prosesser og dermed redusere arbeidsmengden for de ansatte. Når dataene er organisert på riktig måte, blir det enklere å analysere, visualisere og bruke dem i beslutningsprosesser. I stedet for endeløse rapporter og kontroller får

spesialistene tilgang til oppdatert informasjon med noen få klikk eller ved hjelp av LLM -agenter, automatisk i form av ferdige dokumenter og dashbord.

Vi vil få hjelp til datamanipulasjonen av verktøy for kunstig intelligens (AI) og LLM chats. De siste årene har det vært en trend bort fra tradisjonelle CRUD-operasjoner (create, read, update, delete) og over til bruk av store språkmodeller (LLM) for datahåndtering. LLM-er kan tolke naturlig språk og automatisk generere passende databaseforespørsler, noe som forenkler interaksjonen med datahåndteringssystemer (fig. 2.2-3).

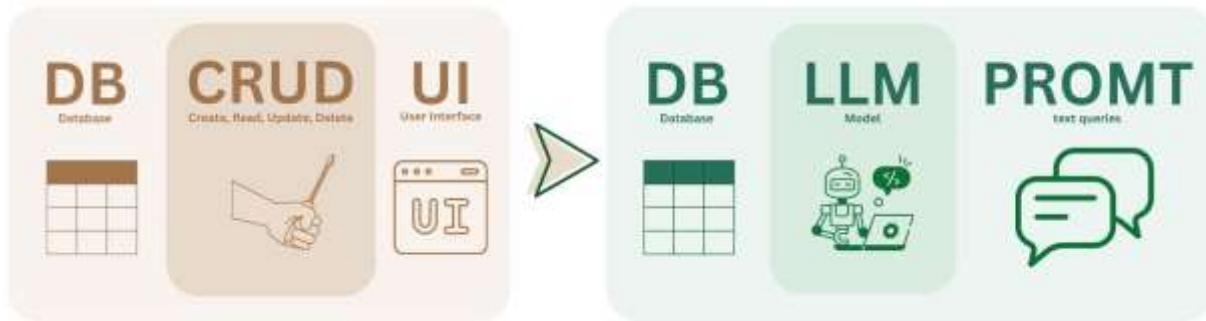


Fig. 2.2-3 AI vil erstatte og integrere lagrings- og databaseløsninger, og gradvis fortrenge tradisjonelle applikasjoner og CRUD -operasjoner.

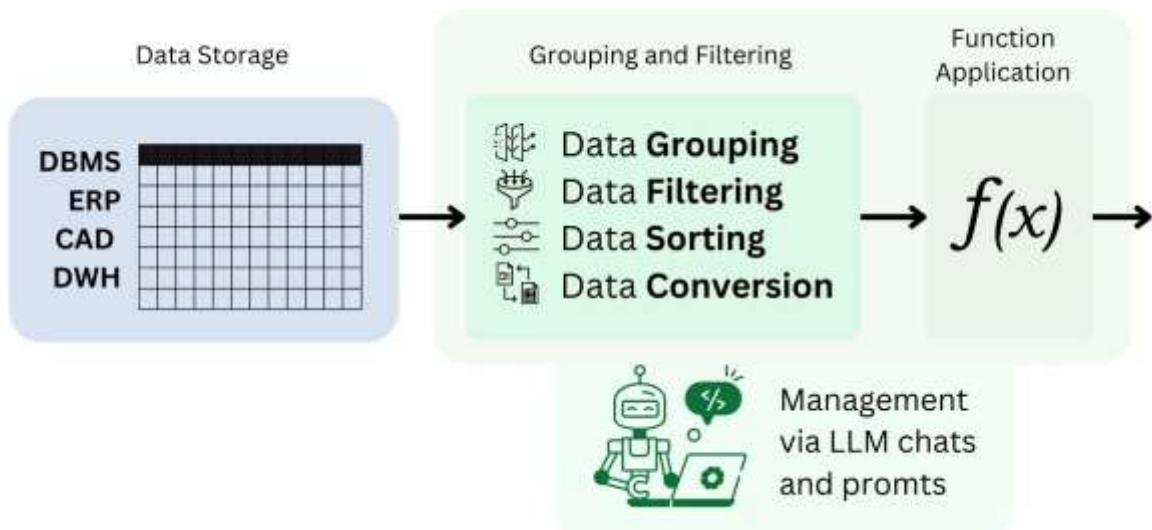
I løpet av de neste 3-6 månedene vil AI skrive 90% av koden, og om 12 måneder kan nesten all koden være generert av AI [47].

- Dario Amodei, administrerende direktør i LLM Anthropic, mars 2025.

Til tross for den raske utviklingen av AI-utviklingsverktøy (f.eks. GitHub Copilot) spiller utviklere fortsatt en nøkkelrolle i denne prosessen i 2025. AI-agenter blir stadig mer nyttige assistenter: De tolker automatisk brukerens -spøringer, genererer SQL- og Pandas-spøringer (mer om dette i de neste kapitlene), eller skriver kode for å analysere data. På denne måten erstatter kunstig intelligens gradvis de tradisjonelle brukergrensesnittene i applikasjoner.

Utbredelsen av modeller for kunstig intelligens, for eksempel språkmodeller, vil drive frem utviklingen av hybridarkitekturen. I stedet for å gå helt bort fra skyløsninger og SaaS-produkter kan vi komme til å se en integrering av skytjenester med lokale datahåndteringssystemer. For eksempel muliggjør føderert læring kraftige AI-modeller uten å måtte flytte sensitive data til skyen. På denne måten kan bedriftene beholde

kontrollen over dataene sine samtidig som de får tilgang til avansert teknologi.



Figur 2.2-4 De grunnleggende operasjonene gruppering, filtrering og sortering, etterfulgt av funksjonsapplikasjon, vil bli håndtert av LLM chats.

Fremtidens byggebransje vil være basert på en kombinasjon av lokale løsninger, skykraft og intelligente modeller som sammen skaper effektive og sikre datahåndteringssystemer. LLM vil gjøre det mulig for brukere uten dyp teknisk kunnskap å samhandle med databaser og datalagre ved å formulere spørsmålene sine på naturlig språk. Vi kommer til å snakke mer om LLM- og AI-agenter og hvordan de fungerer i kapittelet "LLM-agenter og strukturerte dataformater".

Godt organiserte data og enkle, brukervennlige LLM-analyseverktøy vil ikke bare gjøre det enklere å jobbe med informasjon, men også bidra til å minimere feil, øke effektiviteten og automatisere prosesser.

Fra datainnsamling til beslutningstaking: veien mot automatisering

I senere deler av boken vil vi se nærmere på hvordan spesialistene samhandler med hverandre, og hvordan data blir grunnlaget for beslutningstaking, automatisering og driftseffektivitet. Figur 2.2-5 er et eksempel på et diagram som viser rekkefølgen av databehandlingstrinnene i en datasentrisk tilnærming. Dette diagrammet illustrerer Continuous Improvement Pipeline), som vi kommer til å gå nærmere inn på senere i boken.

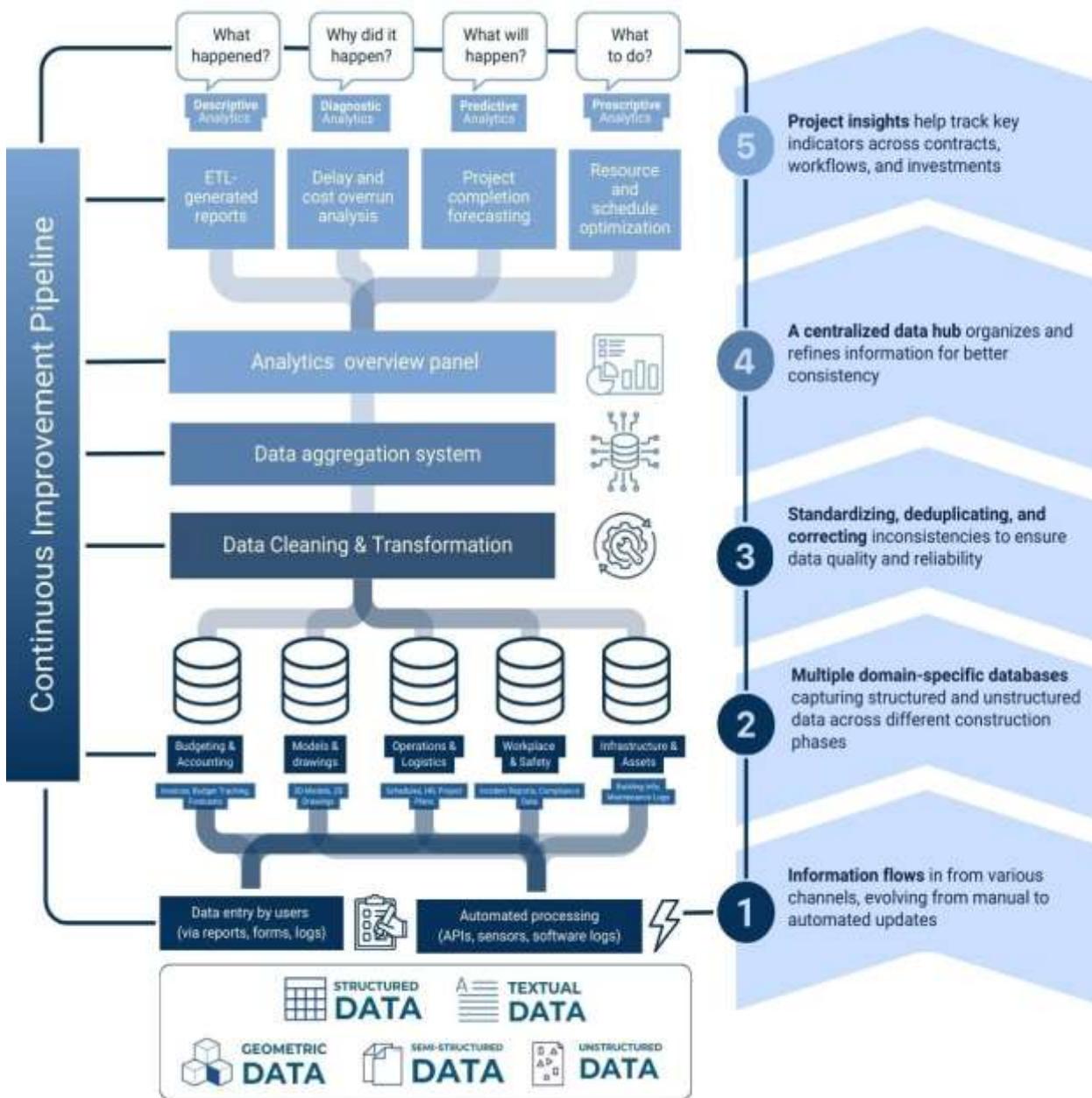


Fig. 2.2-5 Et eksempel på en pipeline for kontinuerlig dataforbedring: databehandlings- og analyseflyten i byggeprosjekter.

Systemet som beskriver forretningsprosessene i en mellomstor bedrift, er bygget på et flernivåprinsipp. Det omfatter datainnsamling, rensing, aggregering, analytisk behandling og beslutningstaking basert på resultatene. Vi vil studere alle disse trinnene senere i boken - både i en teoretisk sammenheng og gjennom praktiske eksempler:

- På det første nivået foregår **dataregistreringen** (fig. 3.1-1). Informasjon mottas både manuelt (gjennom rapporter, skjemaer, logger) og i automatisert form (fra API, sensorer, programvaresystemer). Dataene kan ha ulik struktur: geometriske, tekstlige eller ustrukturerte. På dette stadiet er det behov for standardisering, strukturering og ensretting av informasjonsflyten.

- Neste nivå er **databehandling og -transformasjon**. Det omfatter rensing, fjerning av duplikater, korrigering av feil og klargjøring av informasjon for videre analyse (figur 4.2-5). Dette trinnet er kritisk fordi kvaliteten på analysene avhenger direkte av hvor rene og nøyaktige dataene er.
- **Dataene flyter deretter inn i spesialiserte tabeller, datarammer eller databaser** som er delt inn etter funksjonsområder: budsjetting og regnskap, modeller og tegninger, logistikk, sikkerhet og infrastruktur. Denne inndelingen gjør det enkelt å få tilgang til og analysere informasjon på tvers.
- Dataene **aggregeres og vises i et analytisk dashboard** (showcase). Her brukes deskriptive, diagnostiske, prediktive og preskriptive analysemetoder. Dette bidrar til å besvare nøkkelspørsmål (fig. 1.1-4): hva som skjedde, hvorfor det skjedde, hva som vil skje i fremtiden, og hvilke tiltak som må iverksettes. Systemet kan for eksempel identifisere forsinkelser, forutsi ferdigstillelse av prosjekter eller optimalisere ressurser.
- Det siste nivået genererer **analytiske konklusjoner og nøkkelindikatorer** som bidrar til å overvåke kontraktsoppfyllelse, styre investeringer og forbedre forretningsprosessene (fig. 7.4-2). Denne informasjonen blir grunnlaget for beslutningstaking og selskapets utviklingsstrategi.

På samme måte går data fra å bli samlet inn til å bli brukt i strategisk ledelse. I de neste delene av boken vil vi se nærmere på hvert trinn, med fokus på datatyper, databehandlingsteknikker, analyseverktøy og eksempler fra virkeligheten på hvordan disse tilnærmingene brukes i byggebransjen.

Neste steg: Gjør kaoset om til et håndterbart system

I denne delen utforsket vi utfordringene med informasjonssiloer og så på hvordan overdreven systemkompleksitet påvirker virksomhetens resultater. Vi analyserte overgangen fra den fjerde industrielle revolusjon til den femte, der data snarere enn applikasjoner står i sentrum. Vi så hvordan siloformede informasjonssystemer skaper barrierer for kunnskapsdeling, og hvordan den vedvarende kompleksiteten i IT-landskapet reduserer produktiviteten og hemmer innovasjon i byggebransjen.

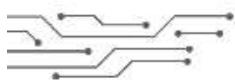
For å oppsummere denne delen er det verdt å fremheve de viktigste praktiske trinnene som vil hjelpe deg med å bruke de diskuterte tilnærmingene i det daglige arbeidet:

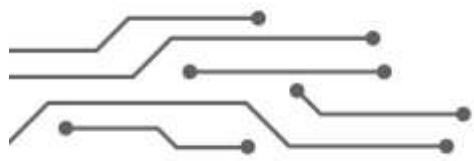
- Visualiser informasjonslandskapet ditt
 - Lag et visuelt kart over datakildene (Miro, Figma, Canva) du jobber med jevnlig
 - Legg til systemene og applikasjonene du bruker i arbeidet ditt på dette kartet
 - Identifiser potensielt dupliserte funksjoner og overflødige løsninger
 - Identifisere kritiske punkter der datatap eller -korrasjon kan oppstå under overføring mellom systemer
- Implementere persontilpassede datahåndteringsrutiner
 - Flytt fokus fra applikasjoner til data som en viktig ressurs i prosessene
 - Dokumentere datakilder og behandlingsmetoder for å sikre åpenhet
 - Utvikle mekanismar for å vurdere og forbedre datakvaliteten
 - Sørg for at data bare legges inn én gang og brukes flere ganger - dette er grunnlaget for en effektiv prosessorganisering
- Fremme en datasentrisk (datadrevet) tilnærming i teamet ditt

-
- Foreslå bruk av standardiserte og enhetlige formater for peer-to-peer-datautveksling
 - Ta jevnlig opp spørsmål knyttet til datakvalitet og tilgjengelighet i teammøter
 - Bli kjent med Open Source alternativer til verktøyene du bruker for å løse problemene dine

Begynn i det små - velg én spesifikk prosess eller ett datasett som er kritisk for arbeidet ditt, og bruk en datasentrisk tilnærming til den, slik at fokus flyttes fra verktøy til data. Ved å oppnå suksess i en enkelt pilot får du ikke bare praktisk erfaring, men også en tydelig demonstrasjon av fordelene med den nye metodikken for teamet ditt. Hvis du har spørsmål i forbindelse med de fleste av disse trinnene, kan du søke avklaring og hjelp fra en oppdatert LLM.

I de neste delene av boken ser vi nærmere på datastrukturering og harmoniseringsteknikker, og vi utforsker praktiske tilnærninger til integrering av heterogen informasjon. Vi vil legge særlig vekt på overgangen fra ulike siloer til enhetlige dataøkosystemer, som spiller en nøkkelrolle i den digitale transformasjonen av byggebransjen.





III DEL

DATARAMMEVERK I FORRETNINGSPROSesser I BYGG- OG ANLEGGSBRANSJEN

I den tredje delen skapes det en omfattende forståelse av typologien for data i bygg- og anleggsbransjen og metoder for effektiv organisering av disse. Vi analyserer kjennetegnene ved og særtrekken ved arbeid med strukturerete, ustrukturerte, halvstrukturerte, tekstlige og geometriske data i forbindelse med byggeprosjekter. Moderne lagringsformater og protokoller for utveksling av informasjon mellom ulike systemer som brukes i bransjen, blir gjennomgått. Praktiske verktøy og teknikker for konvertering av data i flere formater til et enkelt struktureret miljø beskrives, blant annet hvordan man integrerer CAD (BIM) data. Det foreslås metoder for å sikre kvaliteten på data gjennom standardisering og validering, noe som er avgjørende for nøyaktigheten i konstruksjonsberegninger. Praktiske aspekter ved bruk av moderne teknologi (Python Pandas, LLM -modeller) med kodeeksempler for å løse typiske problemer i byggebransjen blir analysert i detalj. Verdien av å opprette et kompetansesenter (CoE) som en organisatorisk struktur for koordinering og standardisering av informasjonshåndteringsmetoder underbygges.

KAPITTEL 3.1.

DATATYPER I BYGG- OG ANLEGGSBRANSJEN

De viktigste datatypene i byggebransjen

I den moderne byggebransjen fylles systemene, applikasjonene og datalagrene til aktivt med informasjon og data av ulike typer og formater (fig. 3.1-1). La oss se nærmere på de viktigste datatypene som utgjør informasjonslandskapet i en moderne bedrift i byggebransjen:

- **Strukturerte** data: Disse dataene har en klar organisasjonsstruktur, f.eks. Excel-regneark og relasjonsdatabaser.
- **Ustrukturerte** data: Dette er informasjon som ikke er organisert i henhold til strenge regler. Eksempler på slike data er tekst, video, bilder og lydopptak.
- **Løst strukturerte** data: Disse dataene befinner seg i en mellomposisjon mellom strukturerte og ustrukturerte data. De inneholder elementer av struktur, men denne strukturen er ikke alltid tydelig eller ofte beskrevet gjennom ulike skjemaer. Eksempler på semistrukturerte data i bygg- og anleggsbransjen er: tekniske spesifikasjoner, prosjektdokumentasjon eller fremdriftsrapporter.
- **Tekstdata:** omfatter alt som stammer fra muntlig og skriftlig kommunikasjon, for eksempel e-poster, utskrifter av møter og avtaler.
- **Geometriske** data: Disse dataene kommer fra CAD-programmene, der spesialister lager geometriske data for prosjektelementer for visualisering, bekreftelse av volumverdier eller kollisjonskontroll.

Det er viktig å merke seg at geometriske data og tekstdata (alfanumeriske data) ikke er en egen kategori, men kan være til stede i alle tre datatypene. Geometriske data kan for eksempel være en del av både strukturerte data (parametriske CAD-formater) og ustrukturerte data (skannede tegninger). Tekstdata kan på samme måte både være organisert i databaser (strukturerte data) og eksistere som dokumenter uten en klar struktur.

Hver type data i et byggefirma er et unikt element i mosaikken av selskapets informasjonsressurser. Fra ustrukturerte data, som bilder fra byggeplasser og lydopptak av møter, til strukturerte poster, inkludert tabeller og databaser, spiller hvert element en viktig rolle i utformingen av selskapets informasjonslandskap.

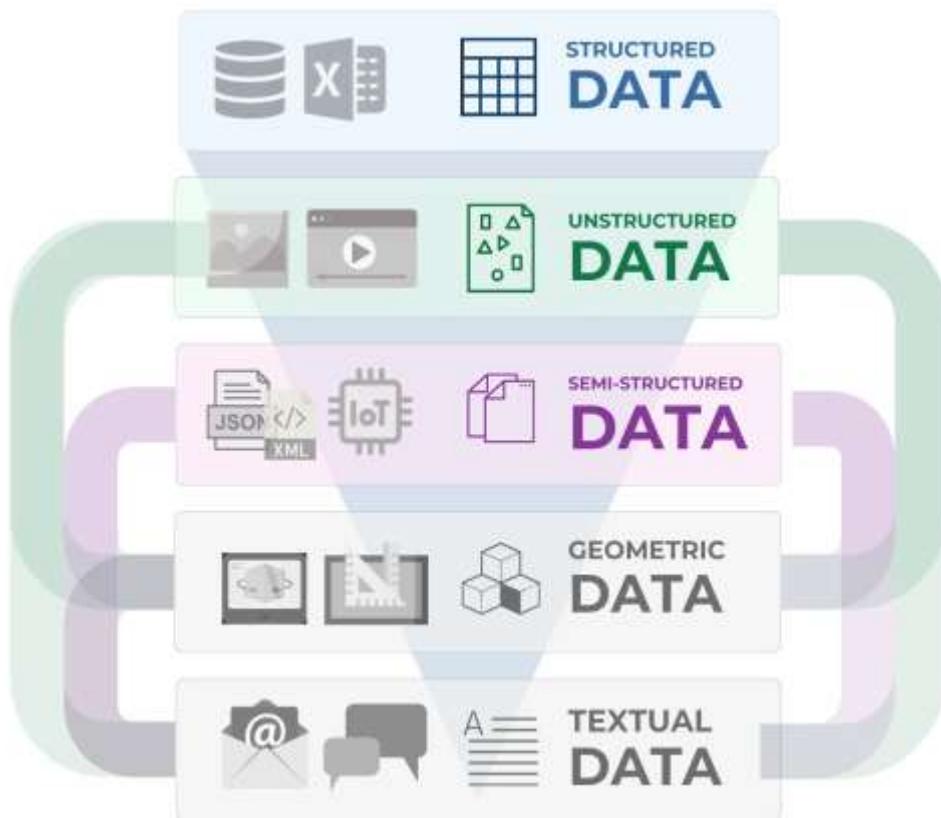


Fig. 3.1-1 Ingeniører og dataansvarlige må lære seg å arbeide med alle typer data som brukes i byggebransjen.

Her er en eksempelliste over noen av systemene og tilhørende datatyper (figur 3.1-2) som brukes i bygg- og anleggsbransjen:

- **ERP** (Enterprise Resource Planning) - håndterer generelt strukturerte data for å administrere virksomhetens ressurser og integrere ulike forretningsprosesser.
- **CAD** (Computer-Aided Design) kombinert med **BIM** (Building Information Modeling) - bruker geometriske og semi-strukturerte data til å designe og modellere byggeprosjekter, noe som sikrer nøyaktighet og konsistens i informasjonen i designfasen.
- **GIS** (Geographic Information Systems) - arbeider med geometriske og strukturerte data for å skape og analysere kartografiske data og romlige relasjoner.
- **RFID** (Radio-Frequency Identification) - bruker halvstrukturerte data til å spore materialer og utstyr på en byggeplass ved hjelp av radiofrekvensidentifikasjon.
- **ECM** (Engineering Content Management) er et system for håndtering av tekniske data og dokumentasjon, inkludert halvstrukturerte og ustrukturerte data som tekniske tegninger og designdokumenter.

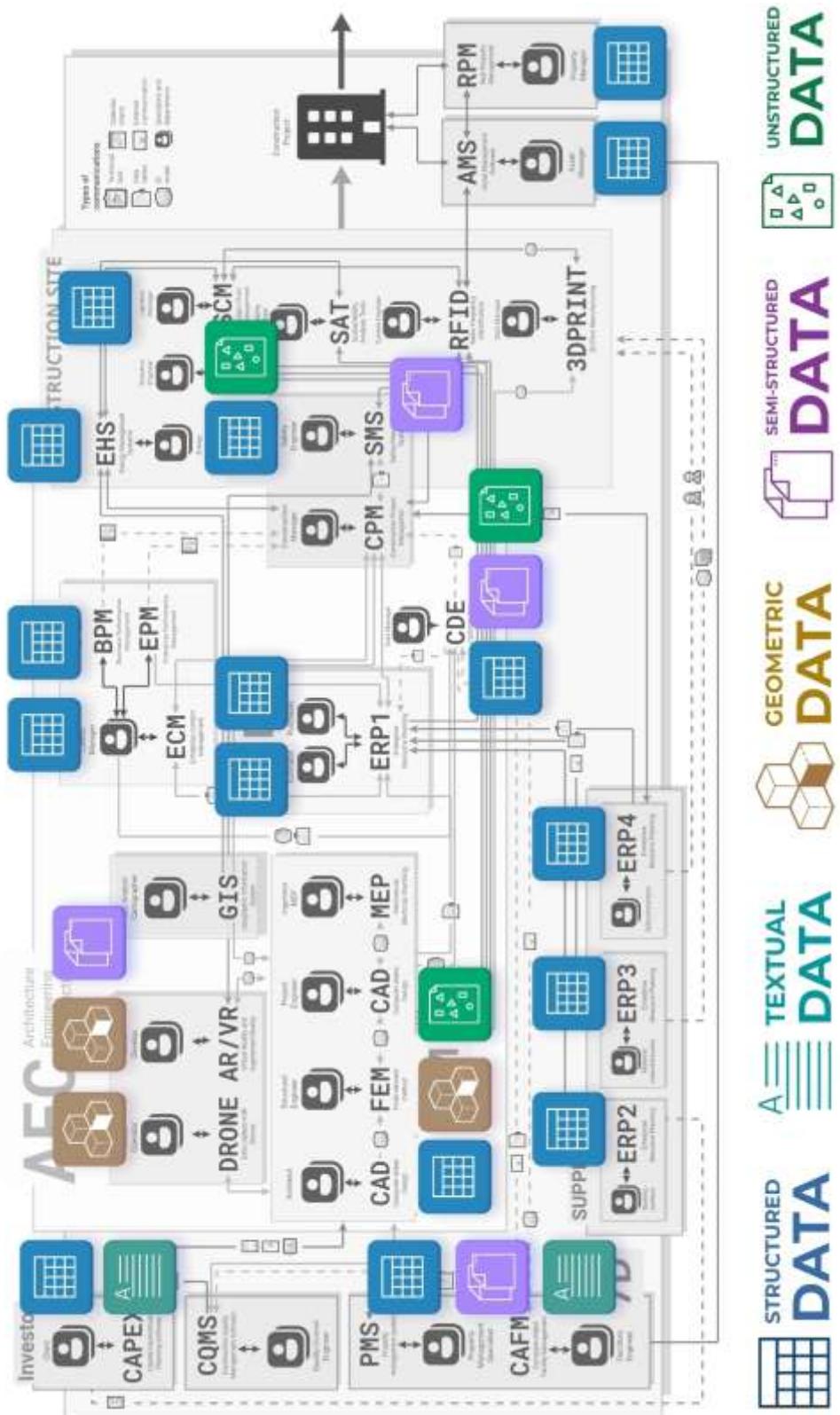


Fig. 3.1-2 Ulike formater og data fyller ulike systemer, noe som krever oversettelse til en form som eigner seg for kompleks integrering.

Disse og mange av selskapets andre systemer håndterer et bredt spekter av data, fra strukturerte tabelldata til komplekse geometriske modeller, og sørger for integrert samarbeid i design-, planleggings- og byggeledelsesprosessene.

I eksempelet på en forenklet dialog (fig. 3.1-3) utveksles ulike typer data mellom spesialistene i byggeprosjektet:

- ⦿ **Arkitekt:** "Med tanke på kundens ønsker har jeg lagt til et sittegruppe på taket. Ta en titt på det nye designet" (geometriske data - modell).
- ⦿ **Bygningsingeniør:** "Prosjektet er mottatt. Jeg beregner bæreevnen til taket på det nye rekreasjonsområdet" (strukturerte og halvstrukturerte data - beregningstabeller).
- ⦿ **Innkjøpssjef:** "Trenger spesifikasjoner og mengder av materialer til rekreasjonsområdet for å organisere innkjøpet" (tekstlige og semi-strukturerte data - lister og spesifikasjoner).
- ⦿ **Helse- og sikkerhetsingeniør:** "Mottatt data om nytt område. Jeg vurderer risikoene og oppdaterer sikkerhetsplanen" (semi-strukturerte data - dokumenter og planer).
- ⦿ **Spesialist på BIM -modellering:** "Gjør endringer i den overordnede prosjektmodellen for å tilpasse arbeidsdokumentasjonen" (geometriske data og semistrukturerte data).
- ⦿ **Prosjektleader:** "Jeg innlemmer den nye rastepllassen i arbeidsplanen. Jeg oppdaterer tidsplaner og ressurser i prosjektstyringssystemet" (strukturerte og halvstrukturerte data - tidsplaner og planer).
- ⦿ **Spesialist på vedlikehold av fasiliteter (FM):** "Jeg forbereder data for fremtidig vedlikehold av rekreasjonsområdet og legger dem inn i eiendomsforvaltningssystemet" (strukturerte og halvstrukturerte data - instruksjoner og vedlikeholdsplaner).

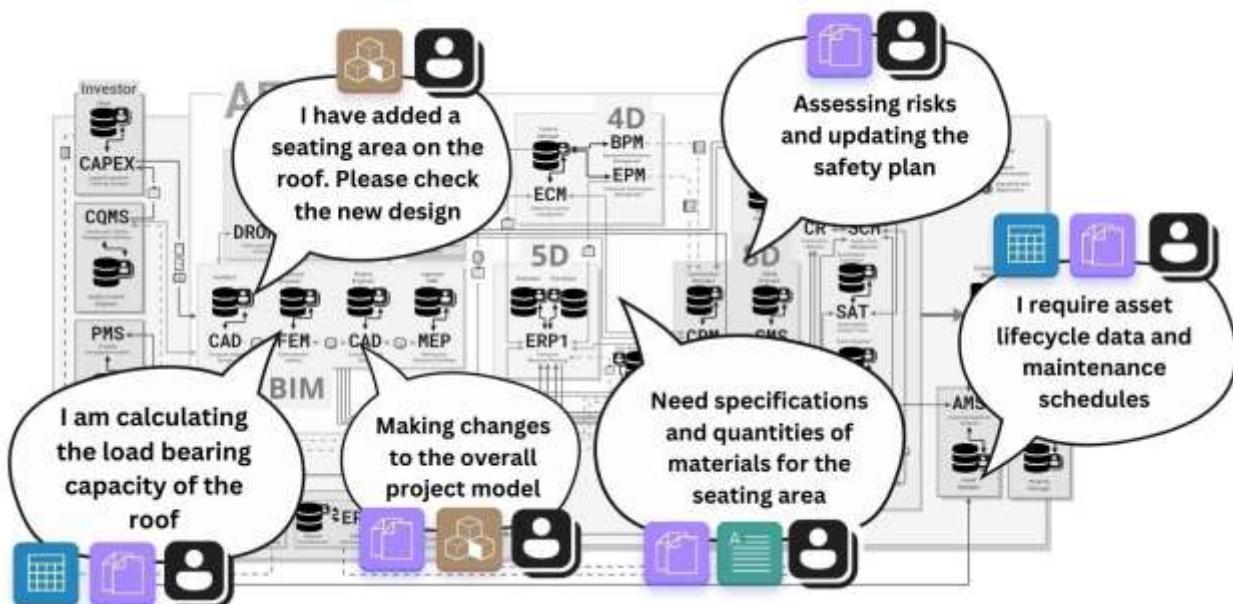


Fig. 3.1-3 Kommunikasjon mellom spesialister foregår både på tekst- og datanivå.

Alle jobber med ulike typer data for å sikre effektivt teamsamarbeid og vellykkede prosjekter. Ved å forstå forskjellene mellom strukturerte, semistrukturerte og ustrukturerte data kan du forstå den unike rollen hver type data spiller i digitale forretningsprosesser. Det er ikke bare viktig å vite at det finnes ulike former for data, men også å forstå hvordan, hvor og hvorfor de brukes.

For ikke så lenge siden virket ideen om å kombinere så ulike data ambisiøs, men vanskelig å realisere. I dag er det allerede en del av daglig praksis. Integrering av data med ulike skjemaer og strukturer er blitt en integrert del av moderne informasjonssystemarkitektur.

I de følgende kapitlene skal vi se nærmere på de viktigste standardene og tilnærmingene som gjør det mulig å kombinere strukturerte, halvstrukturerte og ustukturerte data til én sammenhengende oversikt. Vi vil legge særlig vekt på strukturerte data og relasjonsdatabaser som de viktigste mekanismene for lagring, behandling og analyse av informasjon i byggebransjen.

Strukturerte data

I byggebransjen kommer informasjonen fra mange kilder - tegninger, spesifikasjoner, tidsplaner og rapporter. For å kunne håndtere denne informasjonsflyten på en effektiv måte, må den struktureres. Med strukturerte data kan du organisere informasjonen i en praktisk, lesbar og tilgjengelig form.

Ifølge JB Knowledge's 5th Annual Construction Technology Report [17] sporar og evaluerer 67% av de som jobber med prosjektledelse i bygg- og anleggsbransjen arbeidsutførelsen manuelt eller ved hjelp av regneark.

Noen av de vanligste strukturerte dataformatene er XLSX og CSV. De er mye brukt til lagring, behandling og analyse av informasjon i regneark. I slike regneark presenteres data i form av rader og kolonner, noe som gjør dem enkle å lese, redigere og analysere.

XLSX, et format opprettet av Microsoft, er basert på bruk av XML-strukturer og arkiveres ved hjelp av ZIP-algoritmen. Hovedtrekkene i formatet:

- Støtte for komplekse formler, diagrammer og makroer.
- Mulighet til å lagre data i ulike ark samt formatere informasjon.
- Optimalisert for Microsoft Excel, men kompatibel med andre kontorpakker.

CSV-format er en ren tekstfil der verdiene skiller fra hverandre med komma, semikolon eller andre avgrensningstegn. De viktigste fordelene:

- Universell kompatibilitet med ulike programmer og operativsystemer.
- Enkel import/eksport til databaser og analysesystemer.
- Enkel behandling selv i tekstdredigeringsprogrammer.

CSV støtter imidlertid ikke formler og formatering, så det viktigste bruksområdet er datautveksling mellom systemer og masseoppdatering av informasjon. På grunn av sin allsidighet og plattformuavhengighet har CSV blitt et populært verktøy for dataoverføring i heterogene IT-miljøer.

De to formatene XLSX og CSV fungerer som bindeledd mellom ulike systemer som håndterer strukturerte data (figur 3.1-4). De er spesielt nyttige i oppgaver der lesbarhet manuell redigering og grunnleggende kompatibilitet er viktig.

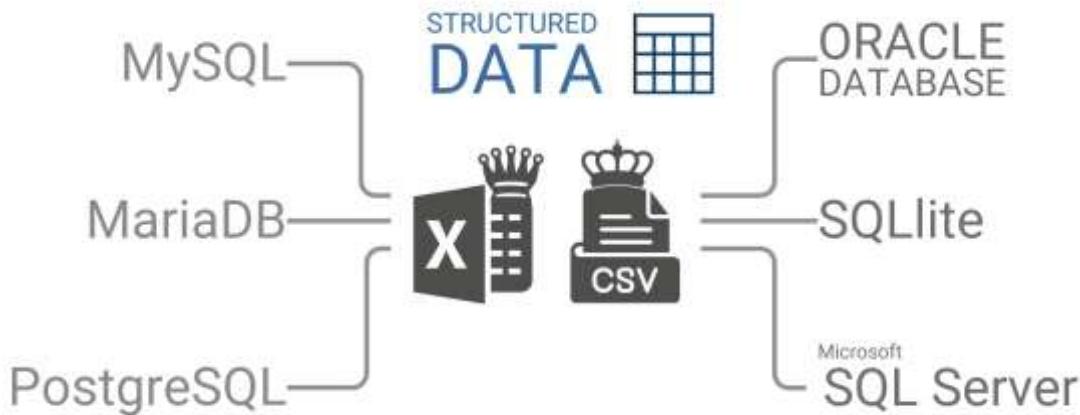


Fig. 3.1-4 XLSX og CSV formatene er bindeleddet mellom ulike systemer som arbeider med strukturerte data.

Plattformuavhengighet gjør CSV til det mest populære formatet for dataoverføring i heterogene IT-miljøer og -systemer.

XLSX og CSV er imidlertid ikke utviklet for databehandling med høy ytelse eller langtidslagring av store datamengder. Til slike formål brukes mer moderne strukturerte formater som Apache Parquet, Apache ORC, Feather, HDF5. Disse formatene vil bli nærmere omtalt i kapittelet Lagring av store datamengder: Analyse av populære formater og deres effektivitet" i del 9 av denne boken.

I praksis brukes Excel med XLSX-format oftere til mindre oppgaver og automatisering av rutineprosesser. Mer komplekse scenarier krever bruk av datahåndteringssystemer som ERP, PMIS CAFM, CPM, SCM og andre (fig. 3.2-1). Disse systemene lagrer strukturerte data som ligger til grunn for organiseringen og styringen av informasjonsflyten i en bedrift.

Moderne informasjonssystemer for datahåndtering i byggebransjen baserer seg på strukturerte data organisert i form av tabeller. For pålitelig, skalerbar og helhetlig håndtering av store informasjonsmengder bruker applikasjons- og systemutviklere relasjonsdatabasesystemer (RDBMS).

Relasjonsdatabaser RDBMS og spørrespråket SQL

For å kunne lagre, behandle og analysere data på en effektiv måte er **relasjonsdatabaser (RDBMS)** datalagringssystemer som organiserer informasjon i tabeller med definerte relasjoner mellom dem.

Data som er organisert i databaser (RDBMS), er ikke bare digital informasjon, de er også grunnlaget for transaksjoner og interaksjon mellom ulike systemer.

Her er noen av de vanligste relasjonsdatabasestyringssystemene (RDBMS) (figur 3.1-5):

- **MySQL** (Open Source) er en av de mest populære RDBMS-ene, som er en del av LAMP-stakken

(Linux, Apache, MySQL, PHP /Perl/Python). Det er mye brukt i webutvikling på grunn av sin enkelhet og høye ytelse.

- **PostgreSQL** (Open Source) er et kraftig objektreasjonssystem kjent for sin pålitelighet og avanserte funksjoner. Det er egnet for komplekse bedriftsløsninger.
- **Microsoft SQL Server** er et kommersielt system fra Microsoft som er mye brukt i bedriftsmiljøer på grunn av integrasjonen med andre bedriftsprodukter og det høye sikkerhetsnivået.
- **Oracle Database** er et av de kraftigste og mest pålitelige DBMS-systemene som brukes i store virksomheter og i virksomhetskritiske applikasjoner.
- **IBM DB2** - rettet mot store selskaper, med høy ytelse og feiltoleranse.
- **SQLite** (Open Source) er en lett, innebygd database, som er ideell for mobile applikasjoner og frittstående systemer som CAD-designprogramvare (BIM).

Populære databasesystemer i byggebransjen - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 og SQLite - arbeider med strukturerte data. Alle disse DBMS-ene er kraftige og fleksible løsninger for håndtering av et bredt spekter av forretningsprosesser og applikasjoner, fra små nettsteder til store bedriftssystemer (fig. 3.2-1).

Ifølge Statista [48] står relasjonsdatabasesystemer (RDBMS) for omtrent 72% av det totale antallet DBMS-er i bruk i 2022.

	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open Source
	5.	5.	5.	MongoDB	Document, Multi-model	Open Source
	6.	7.	9.	Snowflake®	Relational	Commercial
	7.	6.	6.	Redis®	Key-value, Multi-model	Open Source
	8.	8.	7.	Elasticsearch®	Multi-model	Open Source
	9.	9.	8.	IBM Db2	Relational, Multi-model	Commercial
	10.	10.	10.	SQLite	Relational	Open Source
	11.	11.	12.	Apache Cassandra®	Multi-model	Open Source
	12.	12.	11.	Microsoft Access®	Relational	Open Source
	13.	13.	17.	Databricks®	Multi-model	Commercial
	14.	14.	13.	MariaDB	Relational, Multi-model	Open Source
	15.	15.	14.	Splunk	Search engine	Commercial
	16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
	17.	17.	15.	Microsoft Azure SQL	Relational, Multi-model	Commercial

Fig. 3.1-5 Populariteten av å bruke strukturerte databaser (markert med blått) i rangeringen av DBMS (basert på [49]).

Det er ganske enkelt å installere databaser med åpen kildekode - selv uten omfattende tekniske kunnskaper. Open source-systemer, som PostgreSQL, MySQL eller SQLite, er gratis tilgjengelig og fungerer på de fleste operativsystemer: Windows, macOS og Linux. Alt du trenger å gjøre er å gå til prosjektets offisielle nettside, laste ned installasjonsprogrammet og følge instruksjonene. I de fleste tilfeller tar installasjonen ikke mer enn 10-15 minutter. Vi vil modellere og opprette en slik database i den fjerde delen av boken (Fig. 4.3-8).

Hvis bedriften din bruker skytjenester (for eksempel Amazon Web Services, Google Cloud eller Microsoft Azure), kan du distribuere databasen med et par klikk - plattformen tilbyr deg ferdige maler for installasjon. Takket være den åpne koden er slike databaser enkle å tilpasse til dine oppgaver, og et stort brukerfellesskap vil alltid hjelpe deg med å finne en løsning på ethvert problem.

RDBMS er fortsatt grunnlaget for en rekke forretningsapplikasjoner og analyseplattformer (figur 3.1-6) som gjør det mulig for bedrifter å lagre, behandle og analysere data på en effektiv måte - og dermed ta informerte beslutninger i tide.

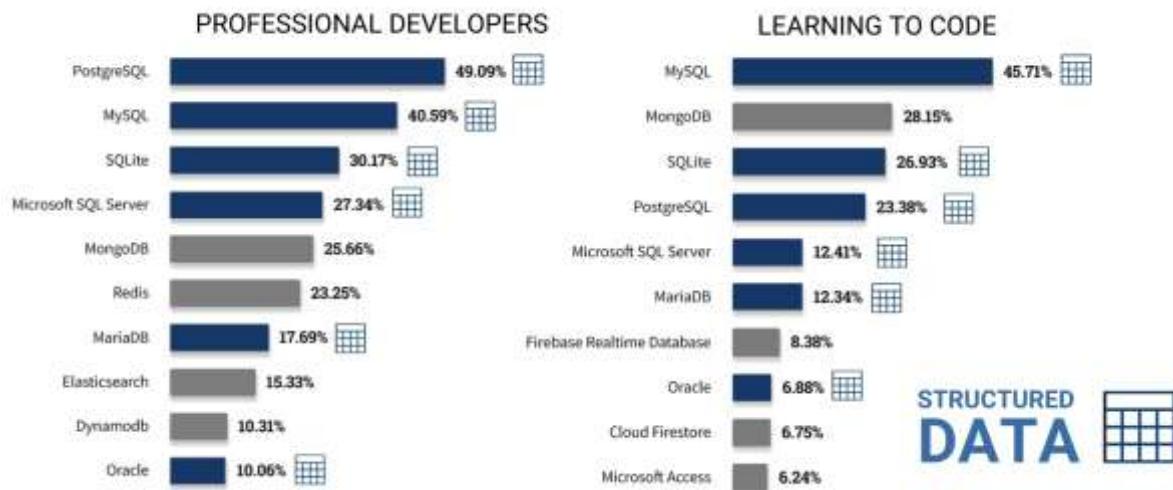


Fig. 3.1-6 Spørreundersøkelse blant utviklere på StackOverflow (det største IT-forumet) om hvilke databaser de brukte i fjor og hvilke de ønsker å bruke neste år (RDBMS-er er uthevet i blått) (basert på [50]).

RDBMS gir pålitelighet, datakonsistens, transaksjonsstøtte og bruker et kraftig spørrespråk - SQL (Structured Query Language), som ofte brukes i analyser og gjør det enkelt å innhente, endre og analysere informasjon som er lagret i databaser. SQL er det viktigste verktøyet for å arbeide med data i relasjonssystemer.

SQL - forespørsler i databaser og nye trender

Den største fordelen med SQL-språket, som ofte brukes i relasjonsdatabaser, i forhold til andre typer informasjonshåndtering (for eksempel ved hjelp av klassiske Excel-regneark), er at det støtter svært store mengder databaser med høy spørringshastighet.

Structured Query Language (SQL) er et spesialisert programmeringsspråk som er utviklet for lagring, behandling og analyse av informasjon i relasjonsdatabaser. SQL brukes til å opprette, administrere og få tilgang til data, slik at du effektivt kan finne, filtrere, kombinere og aggregere informasjon. Det er et viktig verktøy for å få tilgang til data, og det gir en praktisk og formalisert måte å samhandle med informasjonslagre på.

Utviklingen av SEQUEL-SQL-systemer går gjennom viktige produkter og selskaper som Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL og MySQL, og kulminerer i fremveksten av SQLite og MariaDB [51]. SQL gir regnearkfunksjoner som ikke finnes i Excel, noe som gjør datamanipulering mer skalerbar, sikker og enkel å automatisere:

- **Opprette og administrere datastrukturer (DDL):** I SQL kan man opprette, endre og slette tabeller i en database, etablere koblinger mellom dem og definere strukturer for datalagring. I Excel arbeider man derimot med faste ark og celler, uten klart definerte relasjoner mellom ark og datasett.
- **Datamanipulering (DML):** SQL gjør det mulig å legge til, endre, slette og hente ut data i stor skala

og i høy hastighet ved å utføre komplekse spørninger med filtrering, sortering og tabellkoblinger (figur 3.1-7). I Excel krever behandling av store mengder informasjon manuelle handlinger eller spesielle makroer, noe som sinker prosessen og øker sannsynligheten for feil.

- **Tilgangskontroll (DCL):** SQL gjør det mulig å differensiere tilgangsrettighetene til data for ulike brukere, noe som begrenser muligheten til å redigere eller vise informasjon. I Excel er tilgangen derimot enten delt (ved overføring av en fil) eller krever komplekse innstillinger med deling av rettigheter via skytjenester.

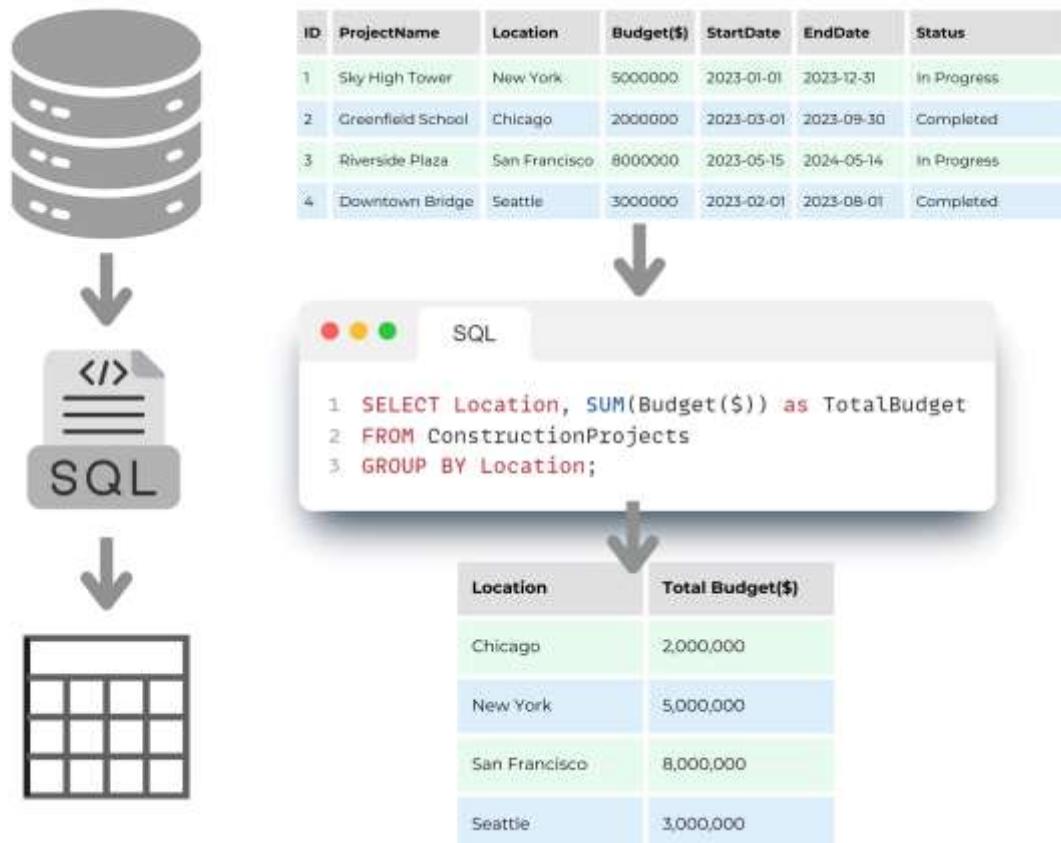


Fig. 3.1-7 Eksempel på DML i SQL: rask behandling, gruppering og aggregering med noen få kodelinjer for automatisk databehandling.

Excel gjør det enklere å jobbe med data med sin visuelle og intuitive struktur. Men etter hvert som datamengden øker, reduseres imidlertid ytelsen i Excel. Excel har også begrensninger på hvor mye data som kan lagres - maksimalt én million rader - og ytelsen blir dårligere lenge før denne grensen nås. Så selv om Excel ser ut til å være å foretrekke for visualisering og manipulering av små datamengder, er SQL bedre egnet til å håndtere store datasett.

Neste trinn i utviklingen av strukturerte data var fremveksten av kolonnedbaser (Columnar Databases), som er et alternativ til tradisjonelle relasjonsdatabaser, spesielt når det gjelder betydelig større datamengder og analytiske beregninger. I motsetning til raddatabaser, der data lagres linje for linje, registrerer kolonnedbaser informasjon kolonnevis. Sammenlignet med klassiske databaser gjør dette dette mulig:

- Reduser lagringsplassen ved å komprimere enhetlige data i kolonner på en effektiv måte.

- Gjør analytiske spørninger raskere, ettersom bare de nødvendige kolonnene leses, ikke hele tabellen.
- Optimaliser Big Data og datalagring, f.eks. Data Lakehouse Architecture.

Vi kommer til å snakke mer om kolonnedbaser, Pandas DataFrame, Apache Parquet, HDF5, samt om å lage Big Data -lagre basert på dem for dataanalyse og -behandling i de følgende kapitlene i denne boken - "DataFrame: et universelt tabellformat" og "Datalagringsformater og arbeid med Apache Parquet: DWH - datavarehus og Data Lakehouse-arkitektur".

Ustrukturerte data

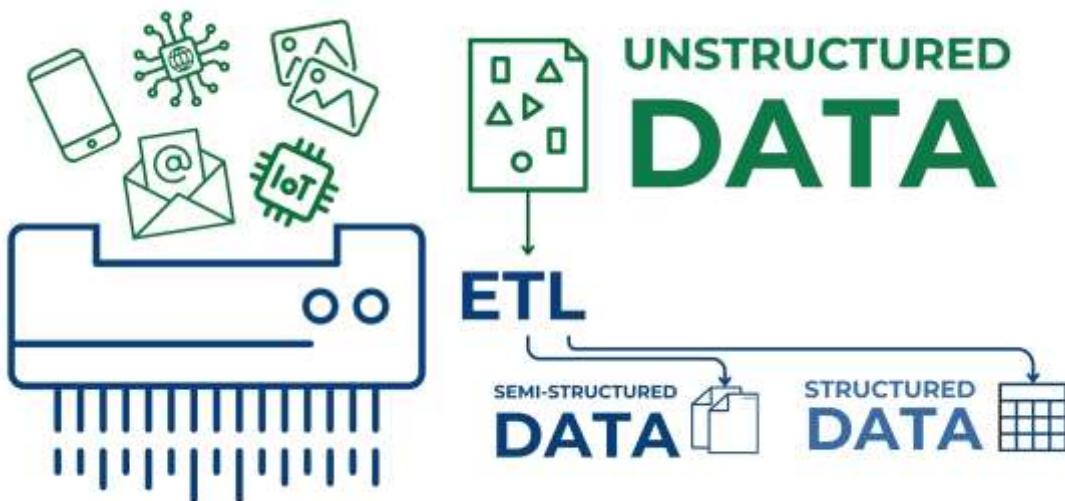
Selv om de fleste data som brukes i applikasjoner og informasjonssystemer er i strukturert form, er det meste av informasjonen som genereres i bygg- og anleggsbransjen, ustrukturert - bilder, videoer, tekstdokumenter, lydopptak og andre former for innhold. Dette gjelder spesielt i bygge-, drifts- og tilsynsfasene, der visuell og teknisk informasjon dominerer.

Ustrukturerte data er informasjon som ikke har noen forhåndsdefinert modell eller struktur, og som ikke er organisert i tradisjonelle rader og kolonner som i databaser eller tabeller.

Generelt sett kan ustrukturerte data klassifiseres i to kategorier:

- Menneskeskapte ustrukturerte data, som omfatter ulike typer menneskeskapt innhold: tekstdokumenter, e-post, bilder, videoer og så videre.
- Ustrukturerte data genereres av maskiner og sensorer, for eksempel loggfiler, GPS-data, resultater fra tingenes internett (IoT) og annen telemetriinformasjon fra for eksempel en byggeplass.

I motsetning til strukturerte data, som enkelt kan organiseres i tabeller og databaser, krever ustrukturerte data flere behandlingstrinn før de kan integreres i informasjonssystemer (fig. 3.1-8). Bruk av teknologi for automatisert innsamling, analyse og omforming av slike data åpner nye muligheter for å forbedre byggeeffektiviteten, redusere feil og minimere påvirkningen fra den menneskelige faktoren.



Figur 3.1-8 Behandlingen av ustrukturerte data begynner med at de omdannes til halvstrukturerte og strukturerde data.

Ustrukturerte data utgjør opptil 80% av all informasjon [52] som fagpersoner i bedrifter kommer i kontakt med, og vi vil derfor gå nærmere inn på typer og behandling av disse dataene med eksempler i de neste kapitlene i boken.

For å gjøre det enklere å diskutere, er tekstdata kategorisert separat. Selv om de er en type ganske ustrukturerte data, krever deres betydning og utbredelse i byggebransjen spesiell oppmerksomhet.

Tekstdata: mellom ustrukturert kaos og strukturert kaos

Tekstdata i byggebransjen dekker et bredt spekter av formater og informasjonstyper, fra papirdokumenter til uformelle kommunikasjonsmetoder som brev, samtaler, arbeidskorrespondanse og muntlige møter på byggeplassen. Alle disse tekstdataene inneholder viktig informasjon for styringen av byggeprosjekter, fra detaljer om designbeslutninger og endringer i planer til diskusjoner om sikkerhetsspørsmål og forhandlinger med entreprenører og kunder (figur 3.1-9).



Fig. 3.1-9 Tekstdata, en av de mest populære informasjonstypene som brukes i kommunikasjonen mellom prosjektdeltakerne.

Tekstlig informasjon kan være både formalisert og ustrukturert. Formaliserte data omfatter Word-

dokumenter (.doc,.docx), PDF-filer, samt tekstfiler med møtereferater (.txt). Uformaliserte data omfatter messenger- og e-postkorrespondanse, møteutskrifter (Teams, Zoom, Google Meet) og lydopptak av diskusjoner (.mp3,.wav) som må konverteres til tekst.

Men mens skriftlige dokumenter som formelle forespørsler, kontraktsvilkår og e-poster som regel allerede har en viss struktur, er muntlig kommunikasjon og arbeidskorrespondanse ofte ustrukturert, noe som gjør det vanskelig å analysere og integrere dem i prosjektstyringssystemer.

Nøkkelen til effektiv håndtering av tekstdata er å konvertere dem til et strukturert format. På den måten kan den bearbeide informasjonen automatisk integreres i eksisterende systemer som allerede arbeider med strukturerte data.



Fig. 3.1-10 Konvertering av tekstlig innhold til strukturerte data.

For å kunne utnytte tekstlig informasjon effektivt må den automatisk konverteres til en strukturert form (figur 3.1-10). Denne prosessen består vanligvis av flere trinn:

- **Tekstgjenkjenning (OCR)** - konvertering av bilder av dokumenter og tegninger til et maskinlesbart format.
- **Tekstanalyse (NLP)** - automatisk identifisering av nøkkelparametere (datoer, beløp og tall som er relevante for prosjektet).
- **Dataklassifisering** - kategorisering av informasjon (økonomi, logistikk, risikostyring).

Etter gjenkjenning og klassifisering kan de allerede strukturerte dataene integreres i databaser og brukes i automatiserte rapporterings- og styringssystemer.

Semistrukturerte og løst strukturerte data

Semistrukturerte data inneholder en viss grad av organisering, men har ikke et strengt skjema eller en streng struktur. Selv om slik informasjon inneholder strukturerte elementer (f.eks. datoer, navn på ansatte og lister over utførte oppgaver), kan presentasjonsformatet variere betydelig fra prosjekt til prosjekt eller til og med fra en ansatt til en annen. Eksempler på slike data er tidslogger, fremdriftsrapporter og tidsplaner, som kan presenteres i en rekke ulike formater.

Semistrukturerte data er enklere å analysere enn ustrukturerte data, men krever ekstra bearbeiding for å kunne integreres i standardiserte prosjektstyringssystemer.

Å arbeide med semistrukturerte data som kjennetegnes av en struktur i stadig endring, byr på betydelige utfordringer. Dette skyldes at variasjonen i datastrukturen krever separate, individuelle tilnærminger til behandling og analyse av hver enkelt kilde til semistrukturerte data.

Men mens det krever mye arbeid å håndtere ustrukturerte data, kan semistrukturerte data behandles med relativt enkle metoder og verktøy.

Svakt strukturerte data er et mer generelt begrep som beskriver data med minimal eller ufullstendig struktur. Oftest dreier det seg om tekstdokumenter, chatter, e-poster, der det finnes noen metadata (f.eks. dato, avsender), men der mesteparten av informasjonen presenteres på en kaotisk måte.

I bygg- og anleggsbransjen finnes løst strukturerte data i en rekke ulike prosesser. De kan for eksempel omfatte:

- Kalkyler og tilbud - tabeller med data om materialer, mengder og kostnader, men uten et enhetlig format.
- Tegninger og tekniske diagrammer - filer i PDF eller DWG, som inneholder tekstkommentarer og metadata, men uten en strengt fast struktur.
- Arbeidsplaner - data fra MS Project, Primavera P6 eller andre systemer, som kan ha en annen eksportstruktur.
- CAD (BIM -modeller) - inneholder elementer av strukturen, men datarepresentasjonen avhenger av programvaren og prosjektstandarden.

Geometriske data, produsert av DAK-systemer, kan kategoriseres på samme måte som semistrukturerte data. Vi vil imidlertid klassifisere geometriske DAK-data (BIM) som en egen type data fordi de, i likhet med tekstdata, ofte kan behandles som en egen datatype i bedriftens prosesser.

Geometriske data og anvendelse

Mens metadata om prosjektelementer nesten alltid lagres i form av tabeller, strukturerte eller løst strukturerte formater, opprettes geometriske data om prosjektelementer i de fleste tilfeller ved hjelp av spesielle CAD-verktøy (fig. 3.1-11), som gjør det mulig å visualisere prosjektelementer i detalj som et sett med linjer (2D) eller geometriske legemer (3D).

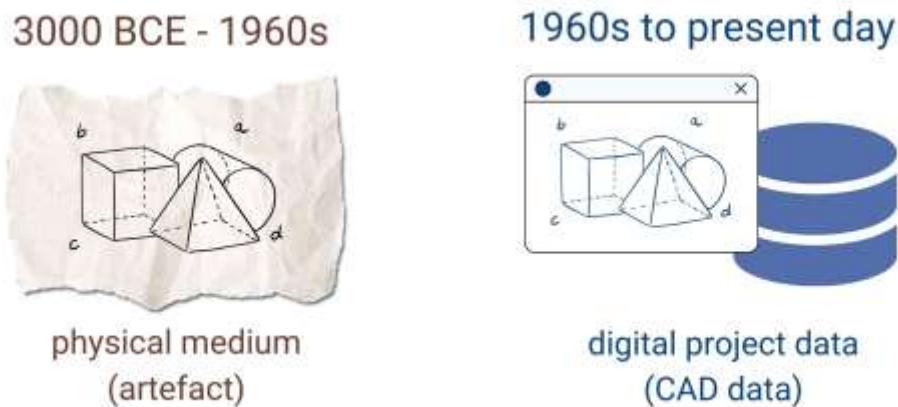


Fig. 3.1-11 CAD verktøy har bidratt til å flytte geometrisk informasjon fra fysiske medier til databaseform.

Når man arbeider med geometriske data innen konstruksjon og arkitektur, kan man identifisere tre hovedbruksområder for geometriske data (figur 3.1-12):

- **Bekrefteelse av volumer:** Geometriske data, generert i CAD-programmer (BIM) ved hjelp av spesielle geometriske kjerner, er nødvendige for automatisk og nøyaktig bestemmelse av volumer og dimensjoner på prosjektelementer. Disse dataene omfatter automatisk beregnede arealer, volumer, lengder og andre viktige attributter som er nødvendige for planlegging, budsjettering og bestilling av ressurser materialer.
- **Visualisering av prosjektet:** Ved eventuelle endringer i prosjektet gjør visualiseringen av elementer det mulig å automatisk generere oppdaterte tegninger i ulike plan. Visualisering av prosjektet i de innledende fasene gir raskere forståelse mellom alle deltakerne, noe som sparer tid og ressurser under byggeprosessen.
- **Kontroll kollisjoner:** I komplekse bygge- og anleggsprosjekter der samspillet mellom flere kategorier av elementer (f.eks. rør og veger) uten "geometriske konflikter" avgjørende, spiller kollisjonskontroll en nøkkelrolle. Ved hjelp av programvaren for kollisjonsdeteksjon kan du proaktivt identifisere potensielle geometriske konflikter mellom prosjektelementer og dermed forhindre kostbare feil under byggeprosessen.

Helt fra begynnelsen av ingeniørdesignkontorene, fra konstruksjonen av de første komplekse strukturene, ga bygningsingeniører geometrisk informasjon i form av tegninger, linjer og flate geometriske elementer (på papyrus, "A0" papir eller i DWG, PDF, PLT-formater), på grunnlag av hvilke formenn og estimatorer (Fig. 3.1-11), i de siste årtusenene, ved hjelp av linjaler og vinkelmålere samlet attributtvolumer eller antall elementer og grupper av elementer. 3.1-11), for de siste årtusenene, ved hjelp av linjaler og transporter, samlet attributive volumer eller mengder av elementer og grupper av elementer.

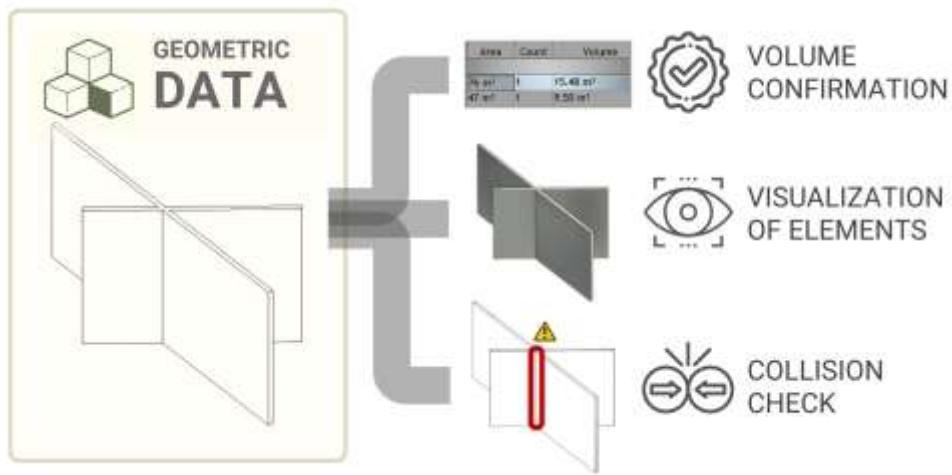


Fig. 3.1-12 Geometri er grunnlaget for å finne volumetriske parametere for elementer, som deretter brukes til å beregne kostnader og tidsbruk for prosjektet.

I dag er denne manuelle og tidkrevende oppgaven fullstendig automatisert takket være fremveksten av volumetrisk modellering i moderne CAD-verktøy (BIM), som gjør det mulig automatisk, ved hjelp av en spesiell geometrisk kjerne, å oppnå volumetriske attributter for et hvilket som helst element uten å måtte beregne volumetriske parametere manuelt.

Moderne DAK-verktøy gjør det også mulig å klassifisere og kategorisere prosjektelelementer slik at man kan laste opp spesifikasjonstabeller fra prosjektdatabasen til bruk i ulike systemer, som for eksempel kostnadsestimering, planlegging eller CO₂-beregning (fig. 3.1-13). Vi går nærmere inn på innhenting av spesifikasjoner, QTO-tabeller og mengder, samt praktiske eksempler i kapittelet "Innhenting av mengder og mengdeberegning".

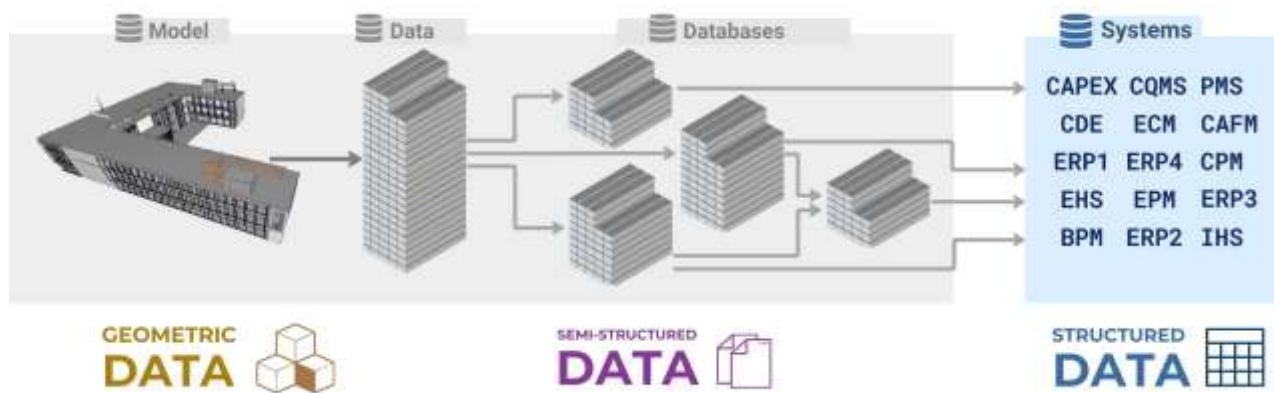


Fig. 3.1-13 DAK-verktøy (BIM) lagrer data i databaser som er utformet for å kunne integreres og samhandle med andre systemer.

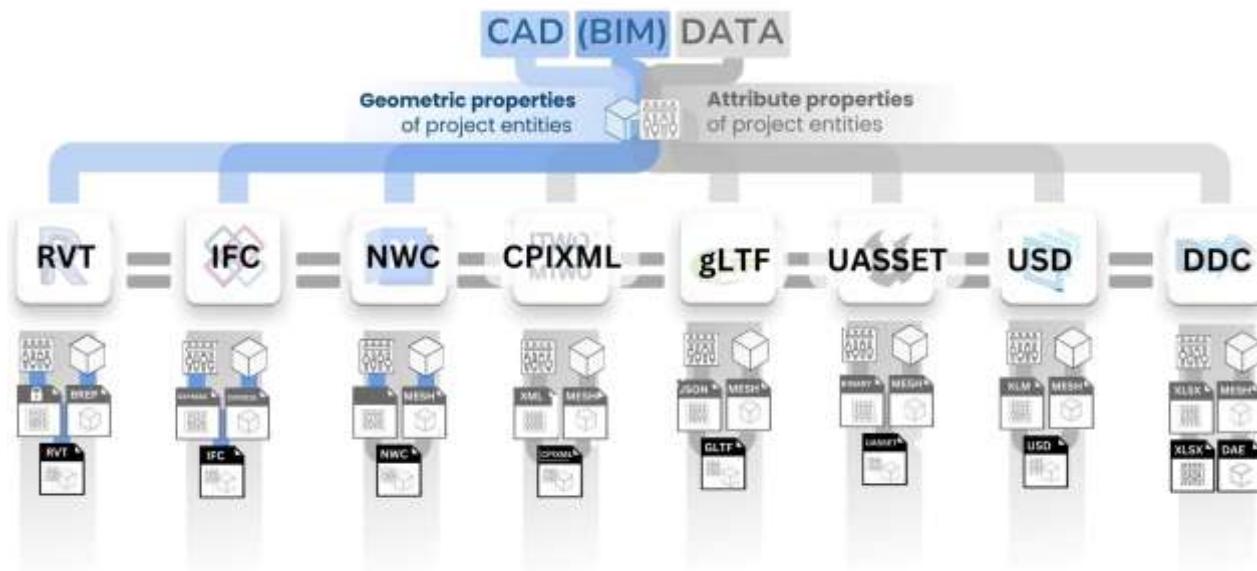
På grunn av de lukkede databasene og formatene som brukes i CAD-miljøer, har geometriske data som skapes i CAD-løsninger, faktisk blitt en egen type informasjon. Den kombinerer både elementgeometri og metainformasjon (strukturert eller halvstrukturert) i spesialiserte filer og formater.

CAD data: fra design til datalagring

Moderne CAD- og BIM-systemer lagrer data i sine egne, ofte proprietære formater: DWG, DXF, RVT, DGN, PLN og andre. Disse formatene støtter både 2D- og 3D-representasjoner av objekter, og bevarer ikke bare geometrien, men også attributtene som er knyttet til objektene. Her er de vanligste:

- **DWG** er et binært filformat som brukes til å lagre todimensjonale (og sjeldnere tredimensjonale) designdata og metadata.
- **DXF** er et tekstformat for utveksling av 2D- og 3D-tegninger mellom CAD-systemer. Det inneholder geometri, lag og attributtdata, og støtter både ASCII og binær representasjon.
- **RVT** er et binært format for lagring av CAD modeller, inkludert 3D-geometri, elementattributter, relasjoner og designparametere.
- **IFC** er et åpent tekstformat for utveksling av konstruksjonsdata mellom CAD (BIM)-systemer. Det omfatter geometri, objektegenskaper og informasjon om relasjonene mellom dem.

I tillegg til disse brukes andre formater: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Selv om de har forskjellig formål og grad av åpenhet (figur 3.1-14), kan de alle representere den samme prosjektinformasjonsmodellen i ulike former. I komplekse prosjekter brukes disse formatene ofte parallelt, fra tegning til koordinering av prosjektmodeller.



Figur 3.1-14 Populære CAD-lagringsformater beskriver geometri ved hjelp av BREP- eller MESH-parametere, supplert med attributtdata.

Alle formatene ovenfor gjør det mulig å lagre data om hvert enkelt element i et byggeprosjekt, og alle formatene inneholder to viktige typer data:

- **Geometriske parametere** - beskriver formen, plasseringen og dimensjonene til et objekt. Geometri og bruken av den vil bli diskutert i detalj i den sjette delen av boken, som er viet CAD-løsninger (BIM);
- **Attributegenskaper** - inneholder ulike opplysninger: materialer, elementtyper, tekniske egenskaper, unike identifikatorer og andre egenskaper som prosjektelementer kan ha.

Attributtdata er spesielt viktige i moderne prosjekter, ettersom de definerer objektenes driftsegenskaper, gir

mulighet for prosjektering og kostnadsberegninger og sørger for ende-til-ende-interaksjon mellom deltakerne i prosjektering, bygging og drift. For eksempel

- For vinduer og dører: konstruksjonstype, glasstype, åpningsretning (fig. 3.2-1).
 - For vegger registreres informasjon om materialer, varmeisolasjon og akustisk ytelse.
 - For tekniske systemer lagres parametere for rørledninger, kanaler, kabeltraseer og deres tilkoblinger.

Disse parameterne kan lagres både i selve CAD-(BIM-)filene og i eksterne databaser - som et resultat av eksport, konvertering eller direkte tilgang til interne CAD-strukturer via reverse engineering-verktøy. Denne tilnærmingen gjør det lettere å integrere designinformasjon med andre bedriftssystemer plattformer.

Reverse engineering i forbindelse med DAK (BIM) er prosessen med å trekke ut og analysere den interne strukturen i en digital modell for å gjenskape logikken, datastrukturen og avhengighetene uten tilgang til de opprinnelige algoritmene eller dokumentasjonen.

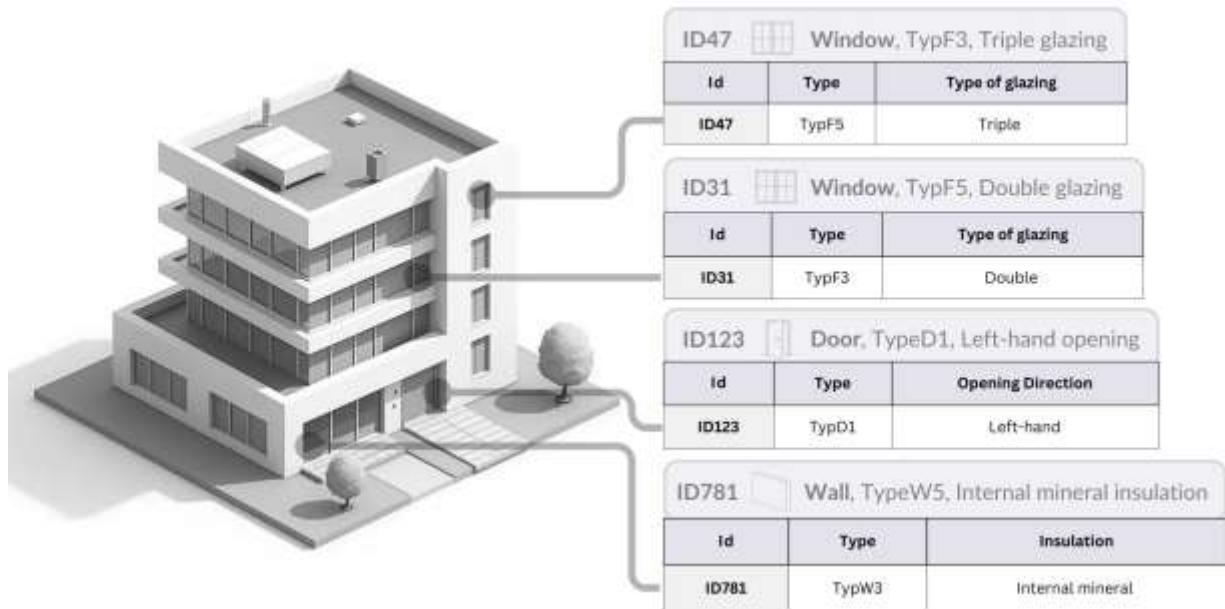


Fig. 3.1-15 Et prosjektelelement inneholder, i tillegg til å beskrive parametrisk eller polygonal geometri, informasjon om parametere og egenskaper til elementer.

Resultatet er at det dannes et unikt sett med parametere og egenskaper rundt hvert element, inkludert både unike egenskaper for hvert objekt (f.eks. identifikator og dimensjoner) og felles attributter for grupper av elementer. Dette gjør det ikke bare mulig å analysere de enkelte elementene i prosjektet, men også å kombinere dem i logiske grupper, som deretter kan brukes av andre spesialister til deres oppgaver og beregninger i systemer og databaser.

En entitet er et konkret eller abstrakt objekt i den virkelige verden som kan identifiseres, beskrives og representeres entydig i form av data.

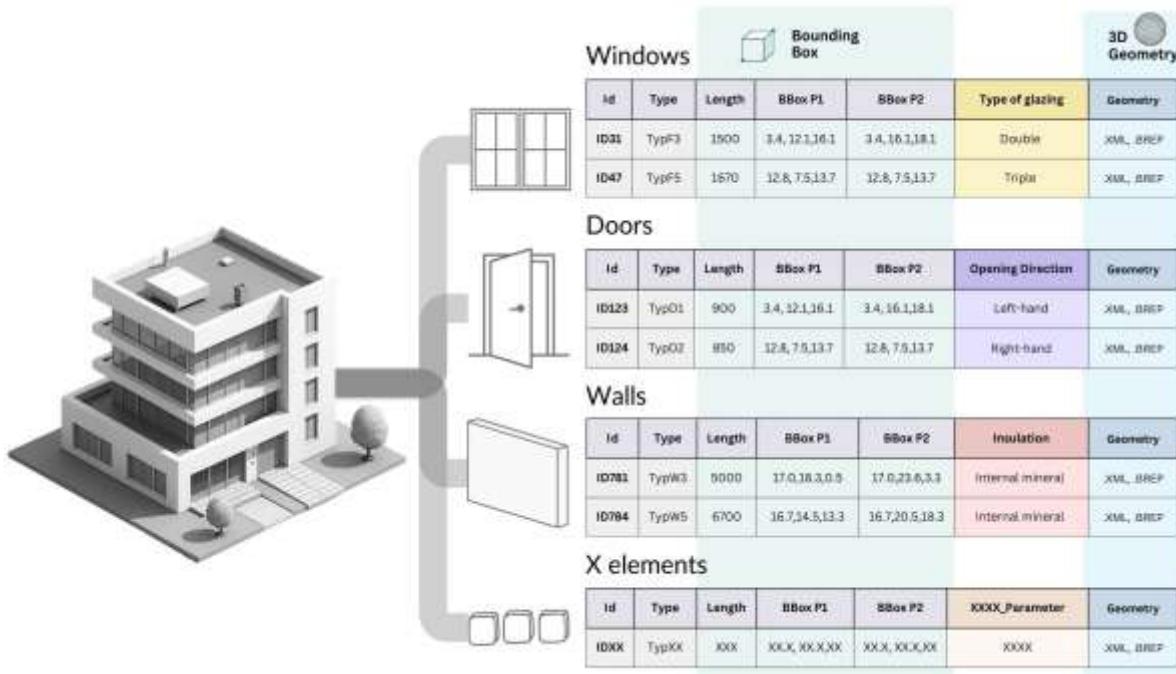


Fig. 3.1-16 Hvert prosjektelement inneholder attributter som enten er lagt inn av konstruktøren eller beregnet i CAD-programmet.

I løpet av de siste tiårene har byggebransjen utviklet mange nye CAD-formater (BIM) som forenkler oppretting, lagring og overføring av data. Disse formatene kan være lukkede eller åpne, tabulære, parametriske eller grafiske. Men mangfoldet og fragmenteringen av formatene gjør datahåndteringen i alle faser av prosjektets livssyklus betydelig mer komplisert. En sammenligningstabell over de viktigste formatene som brukes til informasjonsutveksling i bygg- og anleggsbransjen, er vist i figur 3.1-17 (fullversjon tilgjengelig via QR-kode).

For å løse problemene med interoperabilitet og tilgang til DAK-data, involveres BIM-ledere (BIM) og koordinatorer, som har til oppgave å kontrollere eksport, sjekke datakvalitet og integrere deler av DAK-data (BIM) i andre systemer.

På grunn av formatenes lukkede natur og kompleksitet er det imidlertid vanskelig å automatisere denne prosessen, noe som tvinger spesialister til å utføre mange operasjoner manuelt, uten mulighet til å bygge fullverdige in-line databehandlingsprosesser (pipeline).

The table is a grid structure with many columns and rows. The top row contains the title 'INSTITUTE OF CONSTRUCTION CAD (BIM) DATA STORAGE' and two main categories: 'Geometric properties of project entities' and 'Attribute properties of project entities'. The left side features a vertical column of icons representing different project entities. The body of the table is filled with numerous small colored squares, likely representing data values or types. A QR code is located in the upper right corner of the table area.

Figur 3.1-17 Tabell som sammenligner de viktigste dataformatene som informasjon om prosjektelementer lagres i [53].

For å forstå hvorfor det finnes så mange ulike dataformater, og hvorfor de fleste av dem er lukkede, er det viktig å se nærmere på prosessene som foregår i CAD (BIM) programmer, noe som vil bli utforsket i detalj i den sjette delen av boken.

Et ekstra informasjonslag som ble lagt til geometrien, ble introdusert av utviklere av CAD-systemer i form av BIM-konseptet (Building Information Modeling), et markedsføringsbegrep som har vært aktivt promotert i byggebransjen siden 2002 [54].

Fremveksten av BIM-konseptet (BOM) og bruk av CAD i prosesser

Konseptet bygningsinformasjonsmodellering (BIM), som først ble skissert i Whitepaper fra 2002 [54], stammer fra markedsføringsinitiativene til CAD-programvareprodusentene. Det oppsto som et resultat av markedsføringsinitiativer fra utviklere av CAD-programvare, og var et forsøk på å tilpasse prinsipper som allerede var veletablerte innen maskinteknikk, til behovene i byggebransjen.

Inspirasjonen til BIM kom fra konseptet BOM (Bill of Materials) - Bill of Materials - som har vært mye brukt i industrien siden slutten av 1980-tallet. Innen maskinteknikk gjorde BOM det mulig å koble data fra CAD-systemer med PDM- (Product Data Management), PLM- (Product Lifecycle Management) og ERP-systemer, noe som ga en helhetlig styring av teknisk informasjon gjennom hele produktets livssyklus (fig. 3.1-8).

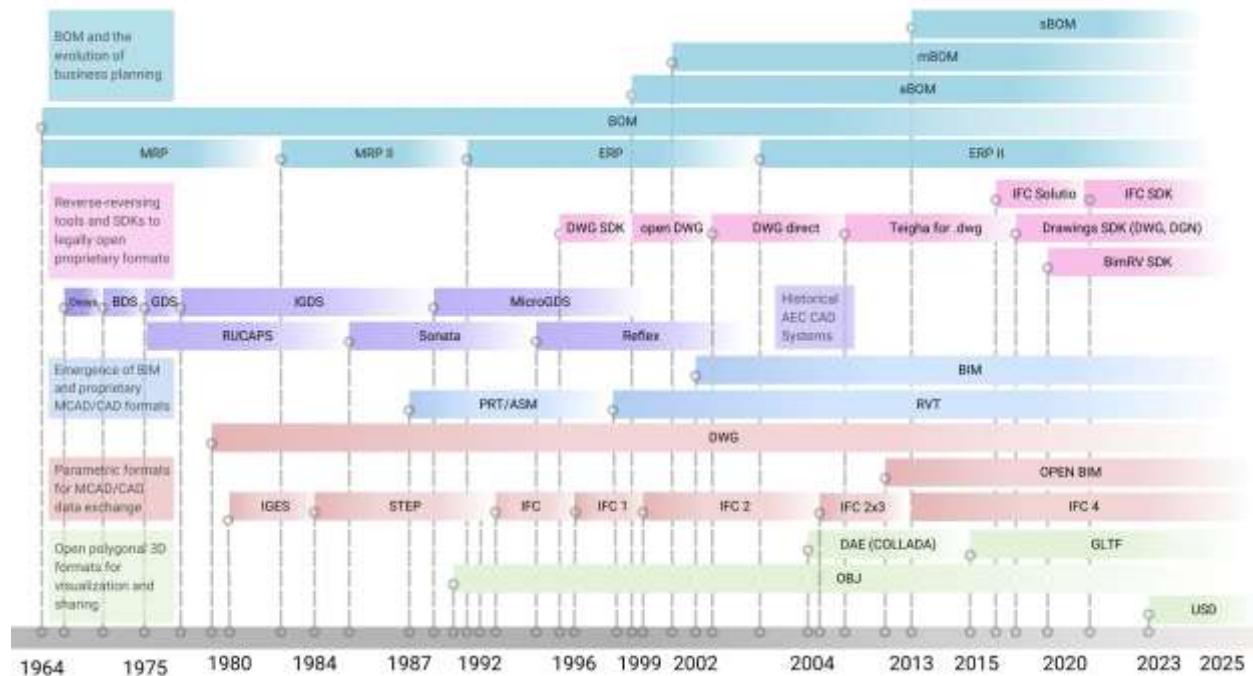


Fig. 3.1-18 Utviklingen av spesifikasjoner (BOM), informasjonsmodellering (BIM)digitale formater i bygge- og anleggsbransjen.

Den moderne utviklingen av BOM-konseptet har ført til fremveksten av et utvidet rammeverk - XBOM (Extended BOM), som ikke bare omfatter produktsammensetning, men også afferdsscenarier, driftskrav, bærekraftparametere og data for prediktiv analyse. XBOM fyller i bunn og grunn samme rolle som BIM i bygg- og anleggsbransjen: Begge tilnærmingene har som mål å gjøre den digitale modellen til en "Single Source of Truth" for alle prosjektdeltakerne gjennom hele prosjektets livssyklus.

En viktig milepål i fremveksten av stykklistene i byggebransjen var introduksjonen av det første parametriske CAD-systemet (MCAD) som var spesielt tilpasset byggebransjen i 2002. Det ble utviklet av teamet som tidligere hadde skapt Pro-E®, et revolusjonerende MCAD-system for maskinteknikk som hadde dukket opp på slutten av 1980-tallet og blitt en industristandard [55].

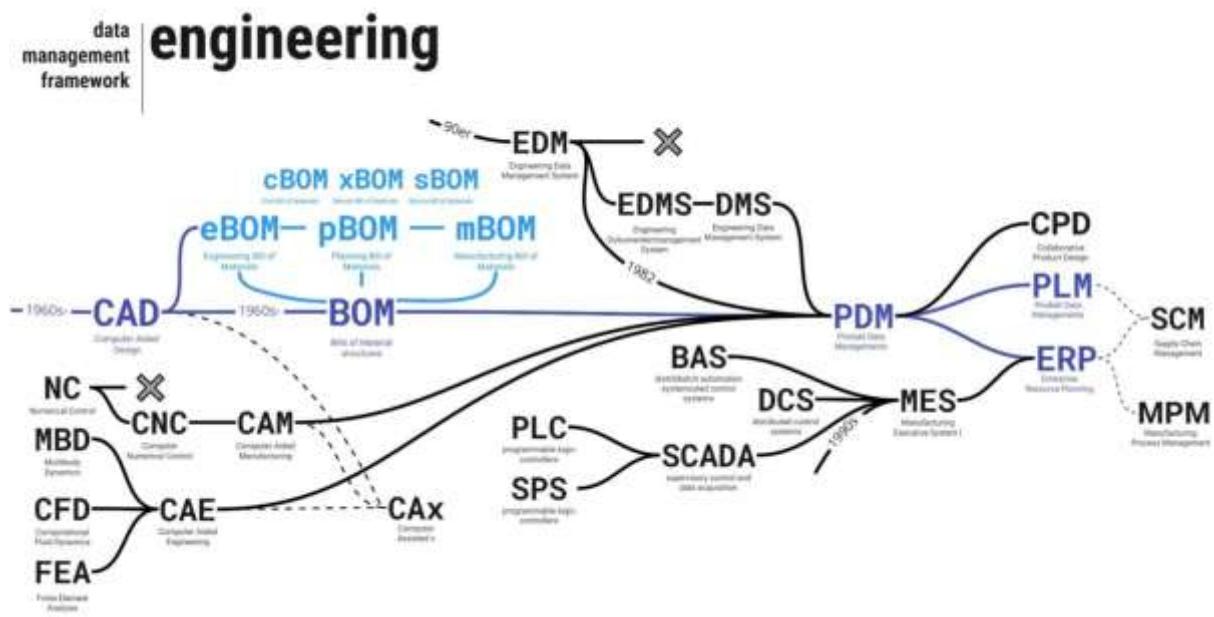
Allerede på slutten av 1980-tallet var målet å eliminere begrensningene [56] ved de daværende CAD-programmene. Hovedmålet var å redusere arbeidsinnsatsen som kreves for å gjøre endringer i parametrene til konstruksjonselementene, og å gjøre det mulig å oppdatere modellen på grunnlag av data utenfor DAK-programmene via en database [57]. Parametrisering skulle spille den viktigste rollen i dette arbeidet: automatisk henting av egenskaper fra databasen og bruk av disse til å oppdatere modellen i DAK-systemene.

Pro-E og konseptet med elementær parametrisk modellering c BOM som ligger til grunn for det, har hatt en betydelig innvirkning på utviklingen av CAD - og MCAD - markedet [58]. Denne modellen har vært i bruk i bransjen i 25 år, og mange moderne systemer har blitt dens konseptuelle etterfølgere.

Målet er å skape et system som er fleksibelt nok til atingeniøren enkelt kan vurdere ulike design. Og kostnadene ved å gjøre endringer i konstruksjonen bør være så nærr null som mulig. Tradisjonell CAD / CAM programvare begrenser på urealistisk vis muligheten til å gjøre billige endringer helt i begynnelsen av designprosessen [59].

- Samuel Heisenberg, grunnlegger av Parametric Technology Corporation®, utvikler av MCAD -produktet Pro-E og lærer av skaperen av et CAD-produkt som bruker RVT-formatet

PDM-, PLM-, MRP- og ERP-systemer har blitt viktige plattformer innen maskinteknikk. De spiller en sentral rolle i data- og prosessstyringen, samler informasjon fra CAx-systemer (CAD, CAM, CAE) og organiserer designaktiviteter basert på produktstrukturen (BOM: eBOM, pBOM, mBOM) (fig. 3.1-18). Denne integrasjonen reduserer antall feil, unngår duplisering av data og sikrer gjennomgående sporbarhet fra design til produksjon.



Figur 3.1-19 Historisk sett oppstod BOM på 1960-tallet som en måte å strukturere data fra CAx-systemer på og overføre dem til kontrollsystemer.

Da en av de ledende leverandørene kjøpte en CAD-løsning som var utviklet av det tidligere Pro-E-teamet og basert på BOM-metoden, ble BIM Whitepaper-serien (2002-2003)[60][61] publisert nesten umiddelbart. Allerede fra midten av 2000-tallet begynte BIM-konseptet å bli aktivt promotert i byggebransjen, noe som økte interessen for parametrisk programvare betraktelig. Populariteten økte så raskt at konstruksjonsgaffelen til maskiningeniør Pro-E - parametrisk CAD fremmet av denne leverandøren - faktisk har fortengt konkurrentene i segmentet for arkitektonisk og strukturell design (fig. 3.1-20). På begynnelsen av 2020-tallet har de de facto konsolidert sin globale dominans i BIM-markedet (CAD) [62].



Fig. 3.1-20 Populariteten til Google-søk (RVT versus IFC): parametrisk CAD laget av det tidligere Pro-E-teamet med BOM-støtte -BIM har vunnet popularitet i nesten alle land i verden.

I løpet av de siste 20 årene har forkortelsen BIM fått en rekke ulike tolkninger, og de mange ulike betydningene har sine røtter i de første markedsføringskonseptene som dukket opp på begynnelsen av 2000-tallet. ISO 19650-standarden, som spilte en viktig rolle i populariseringen av begrepet, sikret faktisk BIM-status som en "vitenskapsbasert" tilnærming til informasjonsstyring. I selve standarden, som omhandler datahåndtering gjennom objektenes livssyklus ved hjelp av BIM, nevnes imidlertid forkortelsen BIM, men den blir aldri klart definert.

Leverandørens opprinnelige nettsted, som publiserte en serie whitepapers om BIM i 2002[60] og 2003[61], reproduksjonen faktisk markedsføringsmateriell om BOM- (Bills of Materials) og PLM- (Product Lifecycle Management) konseptene som tidligere ble brukt i Pro-E-programvaren for maskinteknikk tilbake på 1990-tallet [63].

Bygningsinformasjonsmodellering, en innovativ ny tilnærming til bygningsdesign, konstruksjon og forvaltning som ble introdusert av..... [CAD-leverandørens firmanavn] i 2002, har endret måten bransjefolk over hele verden tenker på hvordan teknologi kan brukes til design, konstruksjon og forvaltning av bygninger.

- BIM Whitepaper, 2003 [61]

Disse tidlige publikasjonene knyttet BIM direkte til konseptet med en sentralisert, integrert database. Som det står i hvitboken fra 2003, er BIM bygningsinformasjonsstyring der alle oppdateringer skjer i ett enkelt arkiv, noe som sikrer at alle tegninger, kutt og spesifikasjoner (BOM - Bills of Materials) er synkroniserte.

BIM beskrives som bygningsinformasjonsforvaltning, der alle oppdateringer og alle endringer skjer i en database. Så enten det dreier seg om skisser, snitt eller plantegninger, er alt alltid koordinert, konsistent og oppdatert.

- CAD-selskapets nettsted vendor with BIM Whitepaper, 2003 [54]

Ideen om å styre design gjennom én integrert database har vært mye diskutert allerede i forskningen på 1980-tallet. For eksempel inneholdt Charles Eastmans BDS-konsept [57] 43 referanser til begrepet "database" (figur 6.1-2). I 2004 var dette antallet nesten halvert til 23 i Whitepaper on BIM fra 2002 [64]. Og på midten av 2000-tallet var temaet databaser praktisk talt forsvunnet fra leverandørenes markedsføringsmateriell og digitaliseringsagendaen generelt.

Selv om det var databasen og tilgangen til den som opprinnelig var tenkt som kjernen i BIM -systemet, har hovedvekten med tiden blitt flyttet til geometri, visualisering og 3D. Den samme leverandøren som i 1994 stod bak IFC-standarden, og som publiserte BIM Whitepaper i 2002, påpekte i Whitepaper på begynnelsen av 2000-tallet eksplisitt begrensningene ved nøytrale formater som IGES, STEP og IFC, og behovet for direkte tilgang til CAD-databaser:

Ulike applikasjoner kan være inkompatible, og data som legges inn på nytt, kan være unøyaktige [...]. Resultatet av tradisjonell datastøttet design [CAD]: høyere kostnader, lengre tid til markedet og lavere produktkvalitet. I dag bruker alle større applikasjoner industristandardgrensesnitt for datautveksling på lavt nivå. Ved å bruke de gamle IGES-standardene eller den nye STEP-standarden [IFC er en de facto og de jure-kopi av STEP/IGES-formatet] til å utveksle data mellom applikasjoner fra ulike leverandører, kan brukerne oppnå en viss datakompatibilitet mellom de beste produktene. Men IGES og STEP fungerer bare på lave nivåer, og de kan ikke utveksle data som er like omfattende som den informasjonen som genereres av dagens ledende applikasjoner [...]. Og selv om disse og andre standarder forbedres nesten daglig, vil de alltid ligge etter dagens leverandørprodukter når det gjelder datarikdom. [...] programmer i en applikasjon må kunne utveksle og bevare datarikdommen uten å måtte ty til nøytrale oversettere som IGES, STEP [IFC] eller PATRAN. I stedet bør rammeprogrammene kunne få direkte tilgang til den underliggende CAD-databasen, slik at detaljene og nøyaktigheten i informasjonen ikke går tapt.

- Whitepaper fra CAD-leverandører (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65].

Allerede på 1980-tallet og tidlig på 2000-tallet ble det viktigste elementet i digital design i CAD-miljøet ansett for å være databasen snarere enn formatfilen eller det nøytrale IFC-formatet. Det ble foreslått at man skulle slutte å bruke oversettere, og at applikasjonene skulle ha direkte tilgang til dataene. På midten av 2020-tallet begynte imidlertid BIM-konseptet i realiteten å ligne en "splitt og hersk"-strategi, der interessene til programvareleverandører som bruker lukkede geometriske kjerner, blir prioritert fremfor utviklingen av åpen informasjonsutveksling.

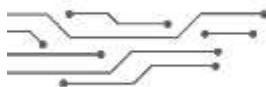
I dag oppfattes BIM som en integrert del av byggebransjen. Men i løpet av de siste to tiårene har løftene om forenklet samarbeid og dataintegrasjon i stor grad ikke blitt innfridd. De fleste løsningene er fortsatt knyttet til lukkede eller nøytrale formater og spesialiserte verktøy. I del 6 av boken "CAD and BIM: Marketing, Reality and the Future of Design Data in Construction" ser vi nærmere på, åpen BIM og IFC, samt spørsmålene om interoperabilitet og geometriske kjerner.

I dag står bransjen overfor en viktig utfordring, nemlig å gå fra den tradisjonelle forståelsen av CAD (BIM) som et modelleringstøyl til å bruke det som en fullverdig database. Dette krever nye måter å arbeide med informasjon på, der man går bort fra avhengighet av lukkede økosystemer og innfører åpne løsninger.

Med utviklingen av verktøy for omvendt konstruksjon som gir tilgang til CAD-databaser, samt utbredelsen av Open Source- og LLM-teknologier, beveger brukere og utviklere i byggebransjen seg i stadig større grad bort fra programvareleverandørenes vase vilkår. I stedet flyttes fokuset over på det som virkelig betyr noe: data (databaser) og prosesser.

Bak de trendy akronymene og visualiseringene skjuler det seg standard datahåndteringspraksis: lagring, overføring og transformasjon - det vil si den klassiske ETL-prosessen (Extract, Transform, Load). Som i andre bransjer krever digitaliseringen av bygg- og anleggsbransjen ikke bare utvekslingsstandarder, men også en tydelig strukturert håndtering av heterogen informasjon.

For å kunne utnytte potensialet i CAD-data (BIM) fullt ut, må bedriftene tenke nytt når det gjelder informasjonsforvaltning. Dette vil uunngåelig føre til et nøkkelement i den digitale transformasjonen - forening, standardisering og meningsfull strukturering av dataene som fagfolk i byggebransjen jobber med til daglig.

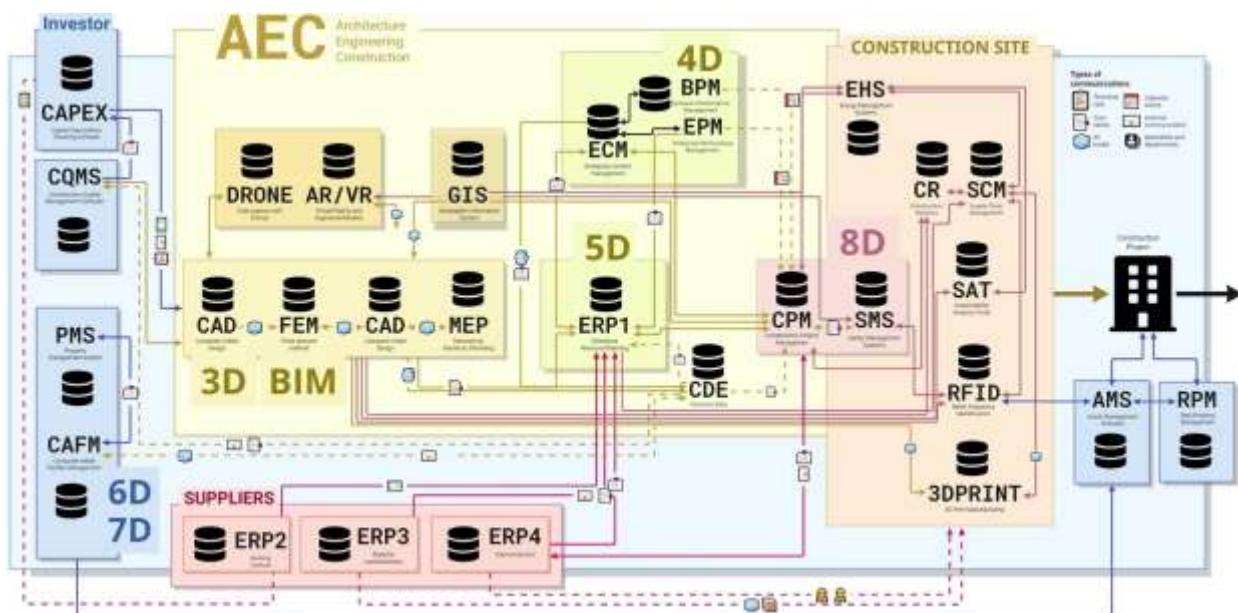


KAPITTEL 3.2.

HARMONISERING OG STRUKTURERING AV DATA

Fyller systemer med data i byggebransjen

Enten det dreier seg om store konsern eller mellomstore bedrifter, er spesialister daglig sysselsatt med å fylle programvaresystemer og databaser med ulike grensesnitt med informasjon i flere formater (fig. 3.2-1), som ved hjelp av ledere må samarbeide med hverandre. Det er dette komplekset av samvirkende systemer og prosesser som til siste og sist skaper inntekter og fortjeneste for bedriften.



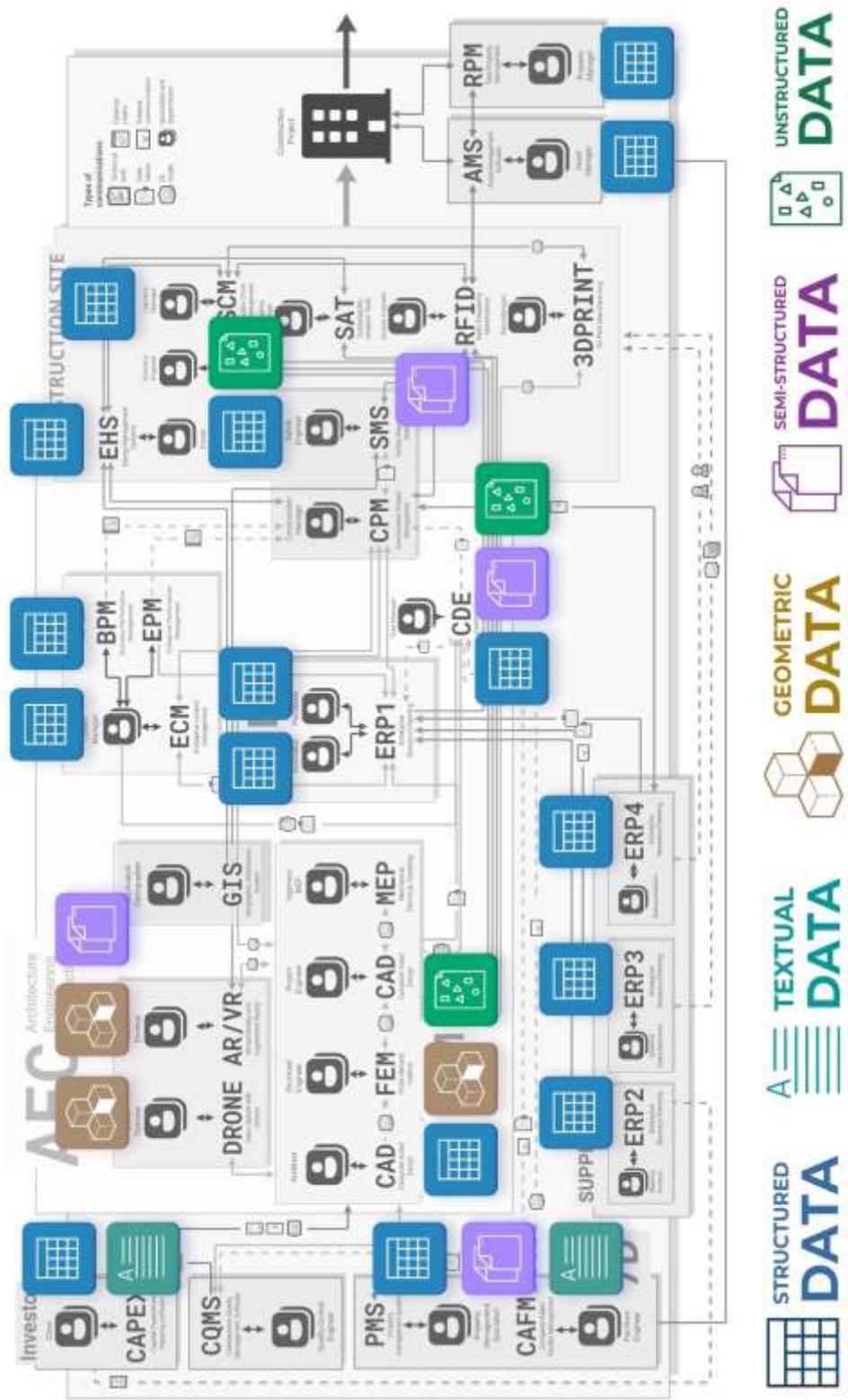
Figur 3.2-1 Så godt som alle systemer eller applikasjoner som brukes i byggebransjen, har en av de populære RDBMS-databasene som kjerne.

Hver av de tidligere nevnte systemkategoriene som brukes i byggebransjen, arbeider med sine egne datatyper som tilsvarer den funksjonelle rollen disse systemene har. For å gå fra det abstrakte nivået til det konkrete, går vi fra datatyper til representasjon av dem som formater og dokumenter.

Til den tidligere listen over systemer (figur 1.2-4) legger vi nå til de spesifikke formatene og dokumenttypene de ofte arbeider med:

■ Investor (CAPEX)

- Finansielle data: budsjetter, utgiftsprøgnosør (strukturerte data).
- Data om markedstrender: markedsanalyser (strukturerte og ustukturerte data).
- Juridiske og kontraktsmessige data: kontrakter (tekstdata).



Figur 3.2-2 Byggebransjen bruker mange systemer med ulike grensesnitt som håndterer forskjellige typer data.

Ledelsessystemer (PMS, CAFM, CQMS)

- Prosjektdata: grafer, oppgaver (strukturerte data).
- Vedlikeholdsdata for anlegget: vedlikeholdsplaner (tekst og semistrukturerte data).
- Kvalitetskontrolldata: standarder, inspeksjonsrapporter (tekstlige og ustrukturerte data).

CAD, FEM og BIM

- Tekniske tegninger: arkitektoniske, strukturelle planer (geometriske data, ustrukturerte data).
- Bygningsmodeller: 3D -modeller, materialdata (geometriske og semistrukturerte data).
- Ingeniørberegninger: belastningsanalyse (strukturerte data).

Ledelsessystemer for byggeplasser (HMS, SCM)

- Sikkerhets- og helsedata: sikkerhetsprotokoller (tekstlige og strukturerte data).
- Data om forsyningskjeden: lagerbeholdninger, bestillinger (strukturerte data).
- Daglige rapporter: arbeidstid, produktivitet (strukturerte data).

Droner, AR/VR, GIS, 3D -printing

- Geodata: topografiske kart (geometriske og strukturerte data).
- Sanntidsdata: video og bilder (ustrukturerte data).
- Modeller for 3D -printing: digitale tegninger (geometriske data).

Ytterligere styringssystemer (4D BPM, 5D ERP1)

- Tids- og kostnadsdata: tidsplaner, estimater (strukturerte data).
- Endringshåndtering: prosjektendringsregistreringer (tekst og strukturerte data).
- Resultatrapportering: indikatorer på suksess (strukturerte data).

Dataintegrasjon og kommunikasjon (CDE, RFID, AMS, RPM)

- Datautveksling: dokumentutveksling, datamodeller (strukturerte og tekstlige data).
- RFID og sporingsdata: logistikk, kapitalforvaltning (strukturerte data).
- Overvåking og kontroll: sensorer på anlegg (strukturerte og ustrukturerte data).

Hvert system i byggebransjen - fra byggeplassadministrasjonssystemer til driftsdatabaser - opererer med sin egen type informasjon: struktureret, tekstlig, geometrisk og annet. "Datalandskapet" som fagfolk må arbeide med til daglig, er ekstremt mangfoldig. En enkel oppramsing av formater avslører imidlertid ikke kompleksiteten i det virkelige arbeidet med informasjon.

I praksis står bedrifter overfor det faktum at data, selv når de hentes ut av systemer, ikke er klare til å brukes "som de er". Dette gjelder spesielt for tekster, bilder, PDF-filer, CAD-filer og andre formater som er vanskelige å analysere med standardverktøy. Derfor er det neste viktige skrittet datatransformasjon - en prosess uten hvilken behandling, analyse, visualisering og beslutningstaking ikke kan automatiseres på en effektiv måte.

Datatransformasjon: det kritiske grunnlaget for moderne forretningsanalyse

I dag står de fleste bedrifter overfor et paradoks: Omrent 80% av de daglige prosessene er fortsatt basert på klassisk strukturerte data - de velkjente Excel-regnearkene og relasjonsdatabaser (RDBMS) [66]. Samtidig er 80% av den nye informasjonen som kommer inn i bedriftenes digitale økosystem, ustrukturert eller løst strukturert (fig. 3.2-3) [52]. Dette omfatter tekst, grafikk, geometri, bilder, CAD -modeller, dokumentasjon i PDF, lyd- og videoopptak, elektronisk korrespondanse og mye mer.

I tillegg fortsetter mengden ustrukturerte data å vokse raskt - den årlige vekstraten anslås til 55-65% [67].

Denne dynamikken gjør det svært vanskelig å integrere ny informasjon i eksisterende forretningsprosesser. Hvis man ignorerer denne strømmen av data i flere formater, oppstår det informasjonshull, og det blir vanskeligere å håndtere hele det digitale miljøet i bedriften.



Fig. 3.2-3 Den årlige veksten i ustrukturerte data skaper utfordringer når det gjelder å integrere strømmende informasjon i forretningsprosesser.

Hvis man ignorerer komplekse, ustrukturerte og uoversiktlige, løst strukturerte data i automatiseringsprosesser, kan det føre til betydelige hull i bedriftens informasjonslandskap. I dagens verden med ukontrollerbare og skredlignende informasjonsbevegelser er det viktig at bedrifter tar i bruk en hybrid tilnærming til datahåndtering som inkluderer effektive metoder for å håndtere alle typer data.

Nøkkelen til effektiv datahåndtering ligger i å organisere, strukturere og klassifisere ulike typer data "Babel" (inkludert ustrukturerte, tekstlige og geometriske formater, til strukturerte eller løst strukturerte data). Denne prosessen forvandler kaotiske datasett til organiserte strukturer som kan integreres i systemer, slik at det blir mulig å ta beslutninger basert på dem (figur 3.2-4).

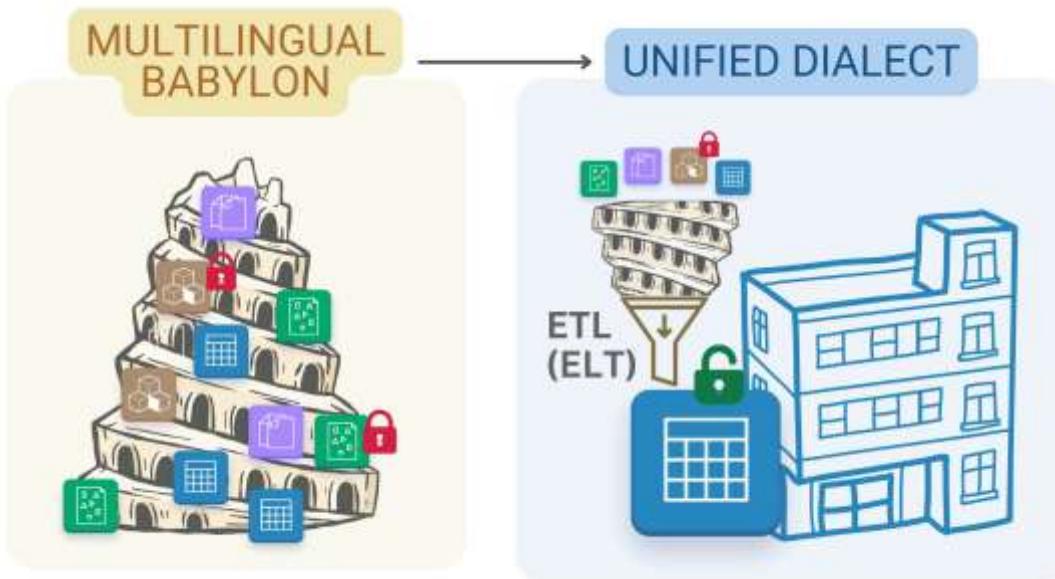


Fig. 3.2-4 Hovedoppgaven til avdelinger for datahåndtering er å oversette "Babylon" av ulike data i flere formater til et strukturert og kategorisert system.

En av de viktigste hindringene for en slik harmonisering er fortsatt den lave graden av interoperabilitet mellom de ulike digitale plattformene - "siloene" vi diskuterte i de foregående kapitlene.

Ifølge rapporten understreker National Institute of Standards and Technology (NIST, USA) [68] at dårlig datakompatibilitet mellom ulike bygningsplattformer fører til tap av informasjon og betydelige ekstrakostnader. Bare i 2002 førte problemer med programvarekompatibilitet til tap på til sammen 15,8 milliarder dollar per år i USA, der to tredjedeler av disse tapene bæres av bygningseiere og -operatører, spesielt under drift og vedlikehold [68]. Studien viser også at standardisering av dataformater kan redusere disse tapene og forbedre effektiviteten gjennom hele anleggets livssyklus.

Ifølge CrowdFlower-studien fra 2016 [69], som omfattet 16 000 dataforskere over hele verden, er hovedproblemet fortsatt "skitne" data og data i flere formater. Ifølge denne studien er den mest verdifulle ressursen ikke de endelige databasene eller maskinlæringsmodellene, men tiden som brukes på å klargjøre informasjon.

Rensing, formatering og organisering tar opptil 60 prosent av tiden til en analytiker og data manager. Nesten en femtedel av tiden går med til å lete etter og samle inn de riktige datasettene, som ofte er gjemt i siloer og utilgjengelige for analyse. Og bare rundt 9 prosent av tiden brukes direkte på modellering, analyse, prediksjoner og hypotesetesting. Resten går med til å kommunisere, visualisere rapportere og lete etter støttende informasjonskilder.

I gjennomsnitt fordeler en leders dataarbeid seg på følgende måte (figur 3.2-5):

- **Rensing og organisering av data (60%):** Rene og strukturerete data kan redusere analytikerens arbeidstid betraktelig og gjøre det raskere å fullføre oppgavene.
- **Datainnsamling (19%):** En av de største utfordringene for fagfolk innen datavitenskap er å finne relevante datasett. Ofte ligger bedriftsdata i kaotisk organiserte "siloer", noe som gjør det

vansklig å få tilgang til den informasjonen de trenger.

- **Modellering/maskinlæring (9%):** Hindres ofte av at kundene har uklare forretningsmål. Mangelen på en klar formålsparagraf kan ødelegge potensialet til selv den beste modellen.
- **Andre oppgaver (5%):** I tillegg til å behandle data må analytikerne drive med forskning, utforske data fra ulike vinkler, kommunisere resultater gjennom visualiseringer og rapporter, og anbefale optimalisering av prosesser og strategier.

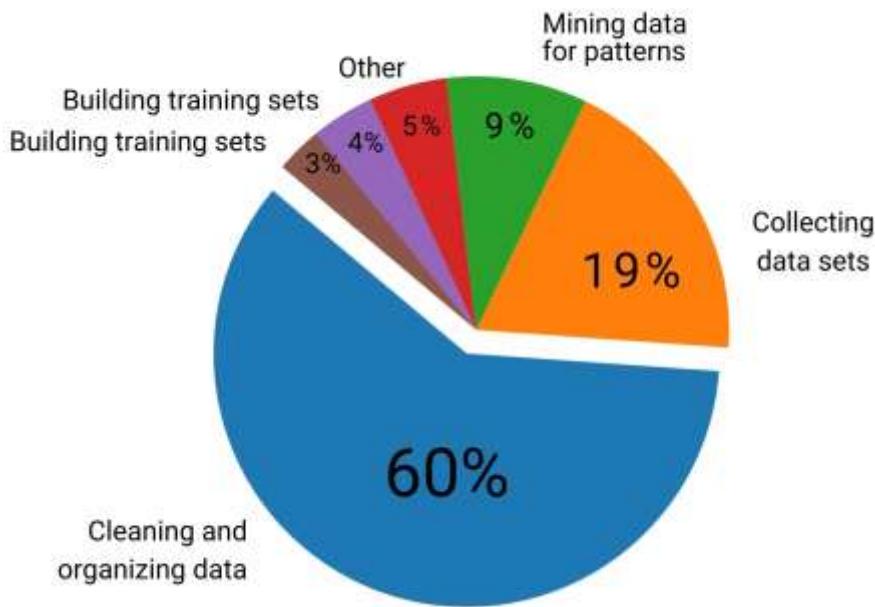


Fig. 3.2-5 Hva dataforvaltere som arbeider med data, bruker mest tid på (basert på [70]).

Disse estimatene støttes av andre studier. Ifølge Xplenty-studien som ble publisert i BizReport i 2015 [71], brukes mellom 50% og 90% av tiden (BI) på å klargjøre data for analyse.

Rensing, validering og organisering av data utgjør et viktig grunnlag for alle data- og analyseprosesser nedstrøms, og tar opptil 90% av tiden til dataforskerne.

Dette møysommelige arbeidet, som er usynlig for sluttbrukeren, er avgjørende. Feil i rådataene forvrenger unngåelig analysene, er villedende og kan føre til kostbare styringsfeil. Derfor er datarensing og standardiseringsprosesser - fra eliminering av duplikater og utfylling av mangler til harmonisering av måleenheter og tilpasning til en felles modell - i ferd med å bli en hjørnesten i moderne digitale strategier.

Grundig omforming, rensing og standardisering av data opptar ikke bare mesteparten av tiden til spesialistene (opptil 80% av arbeidet med data), men er også avgjørende for muligheten til å bruke dem effektivt innenfor rammen av moderne forretningsprosesser. Organisering og rensing av data alene er imidlertid ikke nok til å sikre optimal styring av informasjonsflyten i en bedrift. I løpet av organiserings- og struktureringfasen blir valget av en passende datamodell direkte avgjørende for hvor enkelt og effektivt det er å arbeide med informasjonen i de påfølgende behandlingsfasene.

Siden data og forretningsmål er forskjellige, er det viktig å forstå datamodellenes egenskaper og være i stand til å velge eller skape den riktige strukturen. Avhengig av graden av strukturering og måten relasjonene mellom elementene beskrives på, finnes det tre hovedmodeller: strukturert, løst strukturert og grafisk. Hver

av dem egner seg for ulike oppgaver og har sine egne styrker og svakheter.

Datamodeller: relasjoner i dataene og relasjoner mellom elementer

Data i informasjonssystemer organiseres på ulike måter - avhengig av hvilke oppgaver og krav som stilles til lagring, behandling og overføring av informasjon. Den viktigste forskjellen mellom de ulike typene datamodeller, altså den formen informasjonen lagres i, er graden av strukturering og måten relasjonene mellom elementene beskrives på.

Strukturerte data har et tydelig og repeterbart skjema: de er organisert som tabeller med faste kolonner. Dette formatet gir forutsigbarhet, enkel behandling og effektivitet ved utføring av SQL spørninger, filtrering og aggregering. Eksempler - databaser (RDBMS), Excel, CSV.

Løst strukturerte data tillater fleksibel struktur: ulike elementer kan inneholde ulike sett med attributter og lagres som hierarkier. Eksempler er JSON, XML eller andre dokumentformater. Disse dataene er praktiske når det er nødvendig å modellere nestede objekter og relasjoner mellom dem, men på den annen side kompliserer de dataanalyse og standardisering (fig. 3.2-6).

Data Model	Storage Format	Example
	Relational	CSV, SQL A table of doors in Excel
	Hierarchical	JSON, XML Nested door objects inside a room
	Graph-based	RDF, GraphDB Relationships between building elements

Figur 3.2-6 En datamodell er en logisk struktur som beskriver hvordan data organiseres, lagres og behandles i et system.

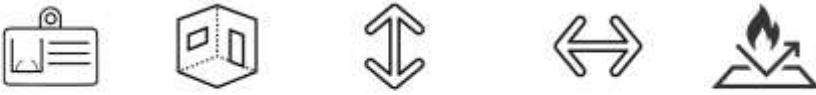
Valget av format avhenger av hvilke mål man har:

- Hvis hastigheten på filtrering og analyse er viktig - relasjonstabeller (SQL, CSV, RDBMS, kolonnedbaser) vil gjøre det.
- Hvis det er behov for en fleksibel struktur, er det bedre å bruke JSON eller XML.
- Hvis dataene har komplekse relasjoner - grafdbaser gir innsyn og skalerbarhet.

I klassiske relasjonsdatabaser (RDBMS) representeres hver enhet (f.eks. en dør) av en rad og dens egenskaper av tabellkolonner. For eksempel kan en tabell med elementer fra kategorien "Dører" inneholde feltene ID, Høyde, Bredde, Brannmotstand og Rom-ID som angir rommet (figur 3.2-7).

I klassiske relasjonsdatabaser (RDBMS) dannes relasjoner i form av tabeller, der hver post representerer et objekt og kolonnene representerer dets parametere. I tabellformatet ser dataene om dører i prosjektet slik

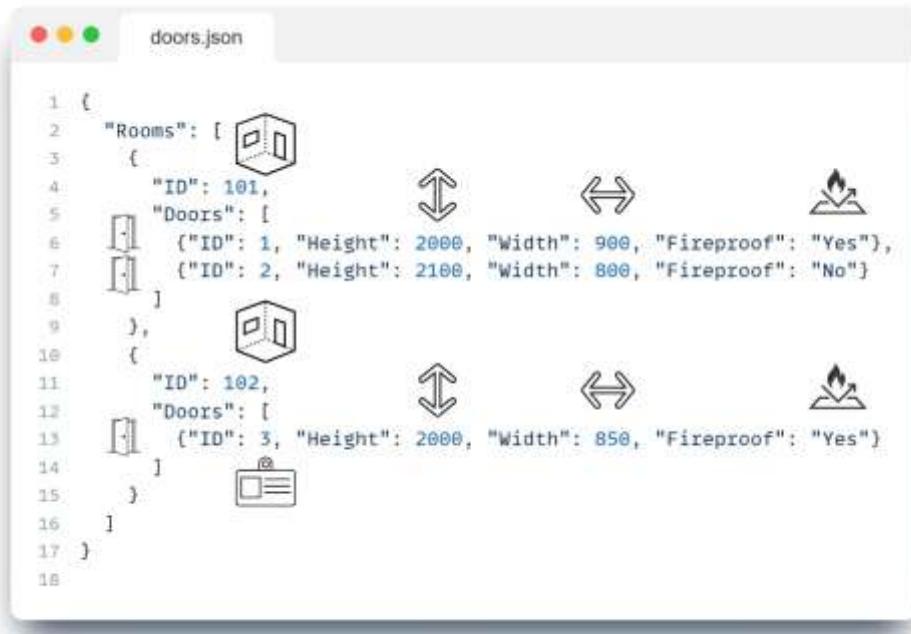
ut, der hver rad representerer et separat element - en dør med sin unike identifikator og attributter, og forbindelsen med rommet utføres gjennom parameteren "Rom-ID".



Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Fig. 3.2-7 Informasjon om de tre elementene i prosjektkategorien "Dører" i tabellstrukturert form.

I løst strukturerede formater som JSON eller XML lagres data i en hierarkisk eller nestet form, der elementer kan inneholde andre objekter og strukturen kan variere. Dette gjør det mulig å modellere komplekse relasjoner mellom elementer. Tilsvarende informasjon om dører i prosjektet, som ble registrert i strukturert form (figur 3.2-7), er representert i et løst strukturert format (JSON) på en slik måte (figur 3.2-8) at de blir nestede objekter innenfor Rooms (Rooms - ID), noe som logisk gjenspeiler hierarkiet.



```

1  {
2    "Rooms": [
3      {
4        "ID": 101,
5        "Doors": [
6          {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7          {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8        ]
9      },
10     {
11       "ID": 102,
12       "Doors": [
13         {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14       ]
15     }
16   ]
17 }
18

```

Fig. 3.2-8 Informasjon om elementene i kategorien "Doors" i prosjektet i JSON-format.

I en grafmodell representeres data som noder (hjørner) og koblinger (kanter) mellom dem. Dette gjør det mulig å visualisere de komplekse relasjonene mellom objekter og deres attributter. Når det gjelder dør- og romdataene i prosjektet, er grafrepresentasjonen av som følger:

- **Noder (knutepunkter)** representerer de viktigste enhetene: rom (rom 101, rom 102) og dører (ID1001, ID1002, ID1003)
- **Ribber (lenker)** viser relasjonene mellom disse enhetene, f.eks. en dørs tilhørighet til et bestemt rom
- **Attributter** mappes til noder og inneholder enhetsegenskaper (høyde, bredde, brannmotstand for dører)



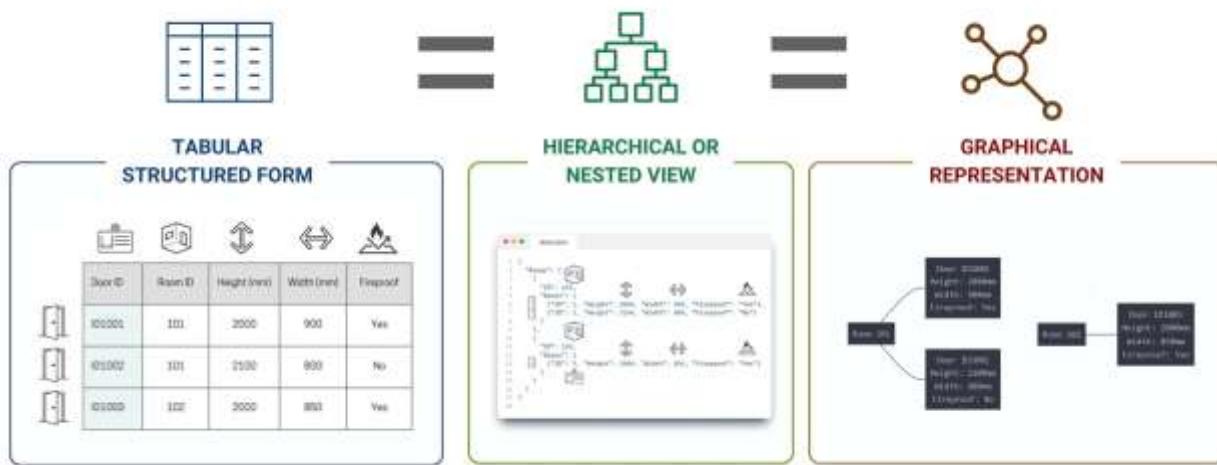
Fig. 3.2-9 Informasjon om prosjektdørens enhet i grafvisning.

I den grafiske datamodellen for dørbeskrivelser er hvert rom og hver dør separate noder. Dørene er knyttet

til rommene gjennom kanter, som indikerer at døren tilhører et bestemt rom. Dørenes attributter (høyde, bredde, branngivende) lagres som egenskaper for de tilhørende nodene. Flere detaljer om grafformatene og hvordan grafsemantikken vokste frem i byggebransjen, vil bli diskutert i kapittelet "Fremveksten av semantikk og ontologi i byggebransjen".

Graf databaser er effektive når det er relasjonene mellom data snarere enn dataene i seg selv som er viktige, for eksempel i anbefalingssystemer, rutingssystemer eller ved modellering av komplekse relasjoner i facility management-prosjekter. Grafformatet gjør det enklere å opprette nye relasjoner ved at nye datatyper kan legges til i grafen uten å endre lagringsstrukturen. Sammenlignet med relasjonstabeller og strukturerte formater gir en graf imidlertid ingen ekstra datakonnektivitet - overføring av todimensjonale databasedata til en graf øker ikke antallet relasjoner og gir ikke ny informasjon.

Dataenes form og skjema bør skreddersys til det spesifikke brukstilfellet og de oppgavene som skal utføres. For å jobbe effektivt i forretningsprosesser er det viktig å bruke de verktøyene og datamodellene som bidrar til å få resultater så raskt og enkelt som mulig.



Figur 3.2-10 Den samme informasjonen om prosjektelementer kan lagres i forskjellige formater ved hjelp av ulike datamodeller.

I dag står de fleste store selskaper overfor problemet med overdreven datakompleksitet. Hver av hundrevis eller tusenvis av applikasjoner bruker sin egen datamodell, noe som skaper overdreven kompleksitet - en individuell modell er ofte flere titalls ganger mer kompleks enn nødvendig, og summen av alle modellene er tusenvis av ganger mer kompleks. Denne overdrevne kompleksiteten vanskeliggjør arbeidet til både utviklere og sluttbrukere.

En slik kompleksitet setter store begrensninger for utvikling og vedlikehold av selskapets systemer. Hvert nytt element i modellen krever ekstra kode, implementering av ny logikk, grundig testing og tilpasning til eksisterende løsninger. Alt dette øker kostnadene og bremser arbeidet til automatiseringsteamet i bedriften, noe som gjør selv enkle oppgaver til kostbare og tidkrevende prosesser.

Kompleksitet påvirker alle nivåer av dataarkitekturen. I relasjonsdatabaser kommer den til uttrykk i det økende antallet tabeller og kolonner, som ofte er overflødige. I objektorienterte systemer øker

kompleksiteten på grunn av de mange klassene og egenskapene som henger sammen. I formater som XML eller JSON kommer kompleksiteten til uttrykk gjennom forvirrende nestede strukturer, unike nøkler og inkonsekvente skjemaer.

Datamodellenes overdrevne kompleksitet gjør ikke bare systemene mindre effektive, men også vanskelig å forstå for slutbrukerne og i fremtiden for store språkmodeller og LLM-agenter. Det er problemet med forståelsen og kompleksiteten i datamodeller og databehandling som reiser spørsmålet om hvordan man kan gjøre data så enkle å bruke at de faktisk begynner å bli nyttige raskt.

Selv når datamodellene er valgt med omhu, reduseres nytten av dem dramatisk hvis tilgangen til dataene er begrenset. Proprietære formater og lukkede plattformer hindrer integrering, kompliserer automatisering og tar bort kontrollen over proprietær informasjon, noe som ikke bare skaper en silo av nye data, men en låst silo som bare er tilgjengelig med leverandørens autorisasjon. For å forstå omfanget av problemet er det viktig å se nærmere på hvordan lukkede systemer påvirker de digitale prosessene i byggebransjen.

Proprietære formater og deres innvirkning på digitale prosesser

En av de største utfordringene som bygg- og anleggsbedrifter står overfor i forbindelse med digitalisering, er begrenset tilgang til data. Dette gjør det vanskelig å integrere systemer, reduserer kvaliteten på informasjonen og gjør det vanskelig å organisere effektive prosesser. Bruken av proprietære formater og lukkede programvareløsninger er ofte årsaken til disse problemene.

Dessverre har mange av programmene som brukes i byggebransjen frem til nå, bare tillatt brukeren å lagre data i proprietære formater eller skylagring, som bare kan nås via strengt begrensede grensesnitt. Og det er ikke uvanlig at disse løsningene er bygget på enda mer lukkede systemer fra større leverandører. Resultatet er at selv de utviklerne som ønsker å tilby mer åpne arkitekturen, blir tvunget til å følge reglene som dikteres av de store leverandørene.

Selv om moderne systemer for styring av byggdata i økende grad støtter åpne formater og standarder (fig. 3.1-5), er CAD- (BIM)-baserte databaser og tilhørende ERP- og CAFM-systemer fortsatt isolerte, proprietære "øyen" i bransjens digitale landskap (fig. 3.2-11).



Figur 3.2-11 Datamaterialets lukkede og proprietære natur skaper barrierer for dataintegrasjon og -tilgang.

Lukkede og monopoliserte formater og protokoller er ikke bare et problem for byggebransjen. I mange sektorer i økonomien har kampen mot lukkede standarder og begrenset tilgang til data startet med treg innovasjon (figur 3.2-12), kunstige etableringshindringer for nye aktører og økende avhengighet av store leverandører. Med den raske veksten i dataenes betydning har konkurransemyndighetene rett og slett ikke tid til å svare på utfordringene fra nye digitale markeder, og resultatet er at lukkede formater og begrenset tilgang til data i realiteten blir digitale "grenser" som informasjonsflyten og veksten [63].

Hvis maskiner produserer alt vi trenger, vil vår situasjon avhenge av hvordan disse varene fordeles. Alle vil bare kunne nyte et liv i velstand hvis rikdommen som produseres av maskiner, deles. Eller så vil folk flest ende opp i dyp fattigdom hvis billeierne lykkes med å drive lobbyvirksomhet mot omfordeling av velstand. Så langt ser det ut til å gå den andre veien, der teknologien fører til stadig større ulikhet [72].

- Stephen Hawking, astrofysiker, 2015

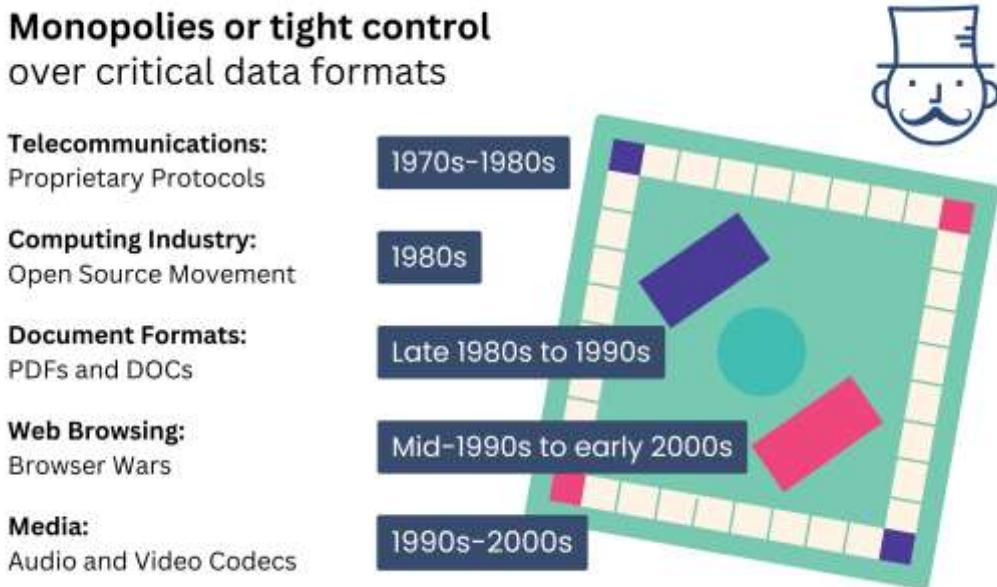


Fig. 3.2-12 Monopol på viktige dataformater og protokoller er ikke bare et problem for byggebransjen.

På grunn av lukket tilgang til databaser programmer, dataforvaltere, dataanalytikere, IT-spesialister og utviklere som lager applikasjoner for datatilgang, -behandling og -automatisering i byggebransjen, står i dag overfor en rekke avhengigheter til programvareleverandører (fig. 3.2-13). Disse avhengighetene i form av flere tilgangslag krever at det lages løsninger med spesialiserte API -tilkoblinger og spesielle verktøy og programvare.

Et API (Application Programming Interface) er et formalisert grensesnitt som gjør det mulig for et program å samhandle med et annet og utveksle data og funksjonalitet uten å måtte gå inn i kildekoden. Et API beskriver hvilke forespørsler et eksternt system kan stille, hvilket format de skal ha, og hvilke svar det vil motta. Det er en standardisert "kontrakt" mellom programvaremодuler.

Det store antallet avhengigheter til lukkede løsninger fører til at hele kodearkitekturen og forretningsprosesslogikken i en bedrift blir en "spaghettiarkitektur" av verktøy som er avhengige av programvareleverandørens policy for å gi kvalitetstilgang til data.

Avhengighet av lukkede løsninger og plattformer fører ikke bare til tap av fleksibilitet, men også til reell forretningsrisiko. Endrede lisensvilkår, stengt tilgang til data, endring av formater eller API-struktur - alt dette kan blokkere kritiske prosesser. Plutselig viser det seg at oppdatering av én tabell krever omarbeiding av en hel blokk med integrasjoner og koblinger (fig. 3.2-13), og enhver storstilt oppdatering av programvaren eller API-leverandøren blir en potensiell trussel mot stabiliteten i hele bedriftens system.

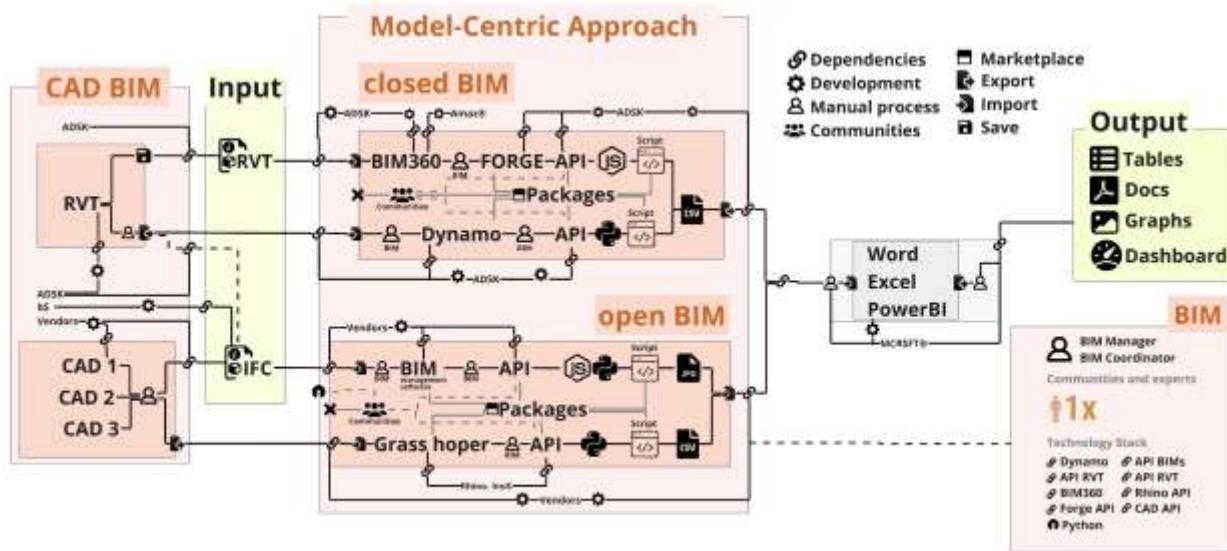


Fig. 3.2-13 Et eksempel på det store antallet avhengigheter i CAD-behandling -data skaper barrierer for dataintegrasjon i økosystemet til byggefirmaer.

Utviklere og systemarkitekter blir under slike forhold tvunget til å jobbe ikke for å forutse, men for å overleve. I stedet for å implementere nye løsninger tilpasser de seg. I stedet for å utvikle, prøver de å opprettholde kompatibilitet. I stedet for å automatisere og fremskynde prosesser bruker de tiden sin på å studere de neste lukkede grensesnittene, API dokumentasjon og endeløse kodeombygginger.

Å arbeide med lukkede formater og systemer er ikke bare en teknisk utfordring - det er en strategisk begrensning. Til tross for de åpenbare mulighetene som ligger i moderne automatisering, AI, LLM og prediktiv analyse, er det mange bedrifter som ikke klarer å realisere sitt fulle potensial. Og barrierene som proprietære formater setter opp (figur 3.2-13), gjør at bedriftene ikke får tilgang til sine egne data. Dette er kanskje ironien i den digitale transformasjonen i byggebransjen.

Datatransparens og åpne systemer er ikke en luksus, men en forutsetning for hastighet og effektivitet. Uten åpenhet blir forretningsprosessene fylt med unødvendig byråkrati, godkjenningskjeder i flere ledd og en økende avhengighet av HiPPO-prinsippet - å ta beslutninger basert på hva den best betalte personen mener.

Likevel er et paradigmeskifte i ferd med å ta form i horisonten. Til tross for at proprietære løsninger dominerer, innser stadig flere selskaper begrensningene i arkitekturen inspirert av den fjerde industrielle revolusjon. I dag går utviklingen i retning av prinsippene i den femte revolusjonen, der data som en strategisk ressurs, åpne grensesnitt (API-er) og ekte interoperabilitet mellom systemer står i sentrum.

Denne overgangen markerer et skifte fra lukkede økosystemer til fleksible, modulære digitale arkitekturen der åpne formater, standarder og transparent datautveksling er nøkkelen.

Åpne formater endrer tilnærmingen til digitalisering

Byggebransjen var en av de siste som tok tak i problemet med lukkede og proprietære data. I motsetning til andre sektorer i økonomien har digitaliseringen gått tregt her. Årsakene til dette er blant annet bransjens tradisjonelt konservative natur, utbredelsen av ulike lokale løsninger og den papirbaserte dokumenthåndteringen som er dypt forankret. I flere tiår har viktige byggeprosesser vært basert på fysiske tegninger, telefonamtaler og usynkroniserte databaser. I denne sammenhengen har lukkede formater lenge blitt oppfattet som normen snarere enn et hinder.

Erfaringer fra andre bransjer viser at fjerning av barrierer for lukkede data fører til økt innovasjon, raskere utvikling og økt konkurranse [73]. Innen vitenskapen kan utveksling av åpne data fremskynde oppdagelser og utvikle internasjonalt samarbeid. Innen medisin kan det gjøre diagnostisering og behandling mer effektiv. Innen programvareteknikk kan det skape økosystemer for samskaping og rask produktforbedring.

Ifølge McKinsey-rapporten "Open Data: Unlock Innovation and Productivity with Information Flow" fra 2013. [74] har åpne data potensial til å frigjøre mellom tre og fem billioner dollar årlig i syv nøkkelbransjer, blant annet bygg og anlegg, transport, helse og energi. Ifølge den samme studien kan desentraliserte dataøkosystemer gjøre det mulig for store byggfirmaer og entreprenører å redusere kostnadene til programvareutvikling og vedlikehold, noe som setter fart på den digitale adopsjonen.

Overgangen til åpne arkitekturen, som for lengst har startet i andre sektorer av økonomien, sprer seg gradvis til byggebransjen. Store selskaper og offentlige oppdragsgivere, og spesielt de finansielle organisasjonene som kontrollerer investeringene i byggeprosjekter, krever i økende grad bruk av åpne data og tilgang til kildekoden for beregninger, kalkyler og applikasjoner. Det forventes ikke lenger at utviklerne bare skal lage digitale løsninger og vise de endelige tallene for et prosjekt - de forventes å være transparente, reproducerte og uavhengige av tredjepartsleverandører av applikasjoner.

Ved å bruke løsninger med åpen kildekode kan kunden være trygg på at selv om eksterne utviklere slutter å samarbeide eller forlater prosjektet, vil det ikke påvirke muligheten til å videreutvikle verktøy og systemer. En av de største fordelene med åpne data er at applikasjonsutviklere ikke lenger er avhengige av bestemte plattformer for å få tilgang til data.

Hvis en bedrift ikke kan gå helt bort fra proprietære løsninger, kan et mulig kompromiss være å bruke reverse engineering-teknikker. Disse juridiske og teknisk forsvarlige metodene gjør det mulig å konvertere lukkede formater til mer tilgjengelige, strukturerte og integrerbare formater. Dette er spesielt viktig når det er behov for å koble seg til eldre systemer eller migrere informasjon fra ett programvarelandskap til et annet.

Et av de beste eksemplene i historien om overgang til åpne formater og bruk av reverse engineering (lovlig hacking av proprietære systemer) i bygg- og anleggsbransjen er kampen for å åpne DWG-formatet, som er mye brukt i systemer for datastøttet design (CAD). I 1998, som svar på monopolet til én programvareleverandør, dannet de 15 andre CAD-leverandørene en ny allianse kalt "Open DWG" for å gi utviklere gratis og uavhengige verktøy for å arbeide med DWG-formatet (de facto-standarden for tegningsoverføring) uten behov for proprietær programvare eller lukkede API-er. Denne hendelsen var et vendepunkt som gjorde det mulig for titusenvis av selskaper å få fri tilgang til det lukkede formatet til en populær CAD-løsning fra slutten av 1980-tallet og frem til i dag, og å skape kompatible løsninger som fremmet konkurransen i CAD-markedet [75]. I dag brukes "Open DWG"-SDK-et, som først ble opprettet i 1996, i nesten alle løsninger der det er mulig å importere, redigere og eksportere DWG-formatet, utenom den offisielle applikasjonen til utvikleren av DWG-formatet.

Andre teknologigiganter er i ferd med å gjennomføre lignende endringer. Microsoft, som en gang var et symbol på en proprietær tilnærming, åpnet kildekoden til .NET Framework, begynte å bruke Linux i Azure-infrastrukturen for skytjenester og kjøpte opp GitHub for å styrke sin posisjon i Open Source-miljøet. [76]. Meta (tidligere Facebook) lanserte AI-modeller med åpen kildekode, for eksempel Llama-serien, for å fremme innovasjon og samarbeid innen utvikling av AI-agenter. Administrerende direktør Mark Zuckerberg forventer at åpen kildekode plattformer vil lede an i teknologiske fremskritt i løpet av det neste tiåret [77].

Open Source er en modell for utvikling og distribusjon av programvare der kildekoden er åpen for fri bruk, studier, modifikasjoner og distribusjon.

Åpne data og åpen kildekode-løsninger er ikke bare i ferd med å bli en trend, men selve grunnlaget for digital bærekraft. De gir bedrifter fleksibilitet, robusthet, kontroll over egne beslutninger og muligheten til å skalere digitale prosesser uten å være avhengig av leverandørens retningslinjer. Og like viktig er det at de gir bedriftene tilbake kontrollen over det 21. århundrets mest verdifulle ressurs - dataene deres.

Paradigmeskifte: Åpen kildekode som slutten på epoken med programvareleverandørers dominans

Byggebransjen gjennomgår et skifte som det ikke er mulig å tjene penger på på vanlig måte. Konseptet med datadrevet, datasentrisk tilnærming og bruk av Open Source-verktøy fører til en revurdering av spillereglene som programvaregiantene på markedet står på.

I motsetning til tidligere teknologitransformasjoner vil denne overgangen ikke bli aktivt fremmet av leverandørene. Paradigmeskiftet truer deres tradisjonelle forretningsmodeller basert på lisensiering, abonnementer og rådgivning. Den nye virkeligheten innebærer ikke et ferdigutviklet produkt eller et betalt abonnement - den krever en omlegging av prosesser og tenkemåter.

For å kunne administrere og utvikle datasenterløsninger basert på åpne teknologier, må bedriftene tenke nytt om interne prosesser. Spesialister fra ulike avdelinger må ikke bare samarbeide, men også tenke nytt om hvordan de jobber sammen.

Det nye paradigmet innebærer bruk av åpne data og Open Source-løsninger, der verktøy basert på kunstig intelligens og store språkmodeller (LLM) snarere enn programmerere vil spille en særlig rolle i å skape programvarekode. Allerede i midten av 2024 er mer enn 25 prosent av ny kode hos Google skapt med kunstig intelligens [78]. I fremtiden vil koding med LLM-er gjøre 80% av arbeidet på bare 20% av tiden (figur 3.2-14).

Ifølge McKinseys 2020-studie [79] erstatter GPU-er i økende grad CPU-er i analysearbeid på grunn av deres høye ytelse og støtte fra moderne Open Source-verktøy. Dette gjør det mulig for selskaper å akselerere databehandlingen uten betydelige investeringer i kostbar programvare eller ansettelse av spesialister.

Ledende konsulentselskaper som McKinsey, PwC og Deloitte understreker den økende betydningen av åpne standarder og åpen kildekode applikasjoner på tvers av bransjer.

Ifølge PwC Open Source Monitor 2019-rapporten [80] bruker 69% av selskaper med 100 eller flere ansatte bevisst Open Source-løsninger. OSS brukes spesielt aktivt i store selskaper: 71% av selskaper med 200-499 ansatte, 78% i kategorien 500-1999 ansatte, og opptil 86% blant selskaper med mer enn 2000 ansatte. Ifølge Synopsys OSSRA 2023-rapporten inneholdt 96% av de analyserte kodebasene komponenter med åpen kildekode [81].

Fremtidens utviklerrolle er ikke å skrive kode manuelt, men å utforme datamodeller, flytarkitekturen og administrere AI-agenter som lager de riktige beregningene på forespørsel. Brukergrensesnittene vil bli minimalistiske, og interaksjonen vil bli dialogbasert. Klassisk programmering vil vike for design og orkestrering av digitale løsninger på høyt nivå (figur 3.2-14). Dagens trender - som lavkodeplattformer (fig. 7.4-6) og LLM-aktiverte økosystemer (fig. 7.4-4) - vil redusere kostnadene ved utvikling og vedlikehold av IT-systemer betraktelig.

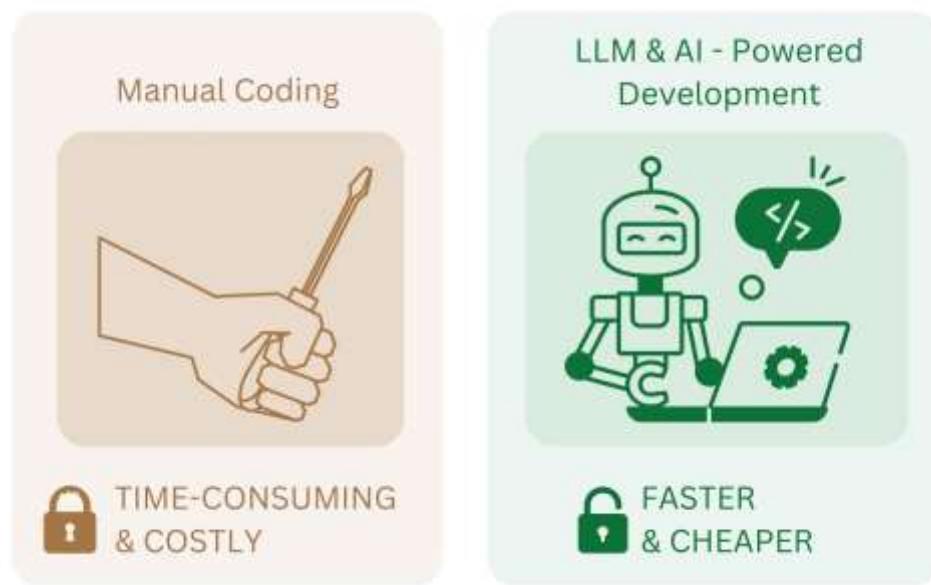


Fig. 3.2-14 Mens applikasjoner i dag lages manuelt av programmerere, vil en betydelig del av koden i fremtiden bli generert av AI- og LLM-baserte løsninger.

Denne overgangen vil være ulik tidligere overganger, og de store programvareleverandørene vil neppe være

katalysatorene.

I Harvard Business School-studien "The Value of Open Source Software" 2024 [40] anslås den totale verdien av programvare med åpen kildekode ut fra to synsvinkler. Hvis vi på den ene siden beregner hvor mye det ville koste å bygge alle eksisterende Open Source-løsninger fra bunnen av, blir beløpet ca. 4,15 milliarder dollar. Hvis vi på den annen side tenker oss at hvert enkelt selskap utvikler sine egne analoger av Open Source-løsninger på egen hånd (noe som skjer overalt), uten å ha tilgang til eksisterende verktøy, vil de totale kostnadene for virksomheten komme opp i kolossale 8,8 billioner dollar - dette er kostnaden for etterspørsel.

Det er ikke vanskelig å gjette seg til at ingen store programvareleverandører er interessert i å krympe et programvaremarked med en potensiell verdi på 8,8 billioner dollar til bare 4,15 milliarder dollar. Dette ville bety å redusere etterspørselsvolumet med mer enn 2000 ganger. En slik transformasjon er rett og slett ulønnsom for leverandører hvis forretningsmodeller er bygget på årevis med kundeavhengighet av lukkede løsninger. Så selskaper som forventer at noen skal tilby dem en praktisk og åpen nøkkelferdig løsning, kan bli skuffet - disse leverandørene kommer rett og slett ikke til å dukke opp.

Overgangen til en åpen digital arkitektur betyr ikke tap av arbeidsplasser eller inntekter. Tvert imot skaper det forutsetninger for fleksible og tilpasningsdyktige forretningsmodeller som på sikt kan fortrenge det tradisjonelle markedet for lisenser og programvare i bokser.

I stedet for å selge lisenser - tjenester, i stedet for lukkede formater - åpne plattformer, i stedet for avhengighet av en leverandør - uavhengighet og muligheten til å bygge løsninger for reelle behov. De som tidligere bare brukte verktøy, vil kunne bli medforfattere av dem. Og de som kan jobbe med data, modeller, scenarier og logikk, vil befinne seg i sentrum av bransjens nye digitale økonomi. Vi kommer til å snakke mer om disse endringene og hvilke nye roller, forretningsmodeller og samarbeidsformer som vokser frem rundt åpne data i den siste, tiende delen av boken.

Løsninger basert på åpne data og åpen kode vil gjøre det mulig for bedrifter å fokusere på effektiviteten i forretningsprosessene i stedet for å slite med utdaterte API-er og integrering av lukkede systemer. En bevisst overgang til åpen arkitektur kan forbedre produktiviteten betydelig og redusere avhengigheten av leverandører.

Overgangen til en ny virkelighet innebærer ikke bare en endring i tilnærmingen til programvareutvikling, men også en revurdering av selve prinsippet om å arbeide med data. I sentrum for denne transformasjonen står ikke koden, men informasjonen: dens struktur, tilgjengelighet og tolkbarhet. Og det er her åpne og strukturerte data kommer i forgrunnen og blir en integrert del av den nye digitale arkitekturen.

Strukturerte åpne data: grunnlaget for digital transformasjon

Mens virksomhetens bærekraft tidligere i stor grad ble bestemt av valg av programvareløsninger og avhengighet av bestemte leverandører, er datakvalitet og evnen til å arbeide effektivt med data den viktigste faktoren i dagens digitale økonomi. Åpen kildekode er en viktig del av det nye teknologiske paradigmet, men potensialet utløses først når dataene er forståelige, organiserte og maskinlesbare. Blant

alle typer datamodeller er strukturerte åpne data i ferd med å bli hjørnestenen i en bærekraftig digital transformasjon.

Den største fordelen med strukturerte åpne data er entydig tolkning og muligheten for automatisert behandling. Dette muliggjør betydelige effektivitetsgevinster både i den enkelte virksomhet og på tvers av organisasjonen.

Ifølge Deloittes rapport "The Data Transfer Process in Enterprise Transformation" [82] er det avgjørende å samarbeide med IT for å håndtere overføringen av strukturerte data. Ifølge den britiske regjeringsrapporten "Data Analytics and AI in Government Project Delivery" (2024) [83] er det å fjerne barrierer for datadeling mellom ulike prosjekter og organisasjoner nøkkelen til å forbedre effektiviteten i prosjektstyringen. I dokumentet understreses det at standardisering av dataformater og innføring av prinsipper for åpne data kan bidra til å unngå duplisering av informasjon, minimere tidstap og forbedre prognosenes nøyaktighet.

For byggebransjen, som tradisjonelt sett har vært preget av en høy grad av fragmentering og et stort mangfold av formater, spiller den strukturerte sammenslåningsprosessen og strukturerte åpne data en avgjørende rolle når det gjelder å skape sammenhengende og håndterbare prosesser (fig. 4.1-14). De gjør det mulig for prosjektdeltakerne å fokusere på å forbedre produktiviteten i stedet for å løse tekniske problemer knyttet til inkompabilitet mellom lukkede plattformer, datamodeller og formater.

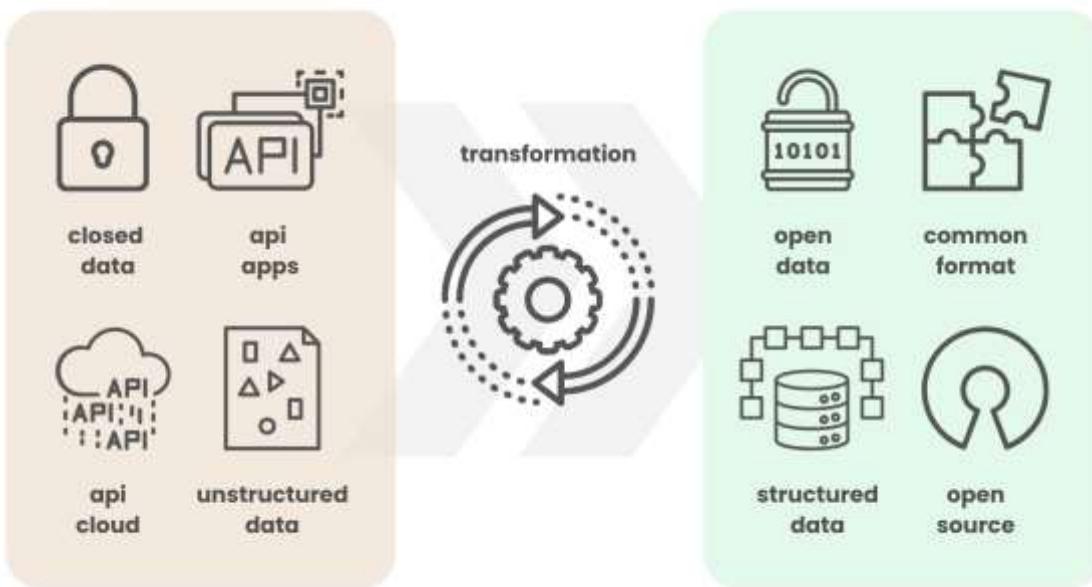


Fig. 3.2-15 Åpne, strukturerte data reduserer avhengigheten av programvareløsninger og plattformer og setter fart på innovasjonen.

Moderne teknologiske verktøy, som vi vil diskutere i detalj senere i boken, gjør det mulig ikke bare å samle inn informasjon, men også å rense den automatisk: eliminere dupliseringer, korrigere feil og normalisere verdier. Dette betyr at analyticere og ingeniører ikke arbeider med ulike dokumenter, men med en organisert kunnskapsbase som er egnet for analyse, automatisering og beslutningstaking.

Gjør det så enkelt som mulig, men ikke enklere.

- Albert Einstein, teoretisk fysiker (sitatets opphavsmann er omstridt [84])

I dag kan de fleste brukergrensesnitt for arbeid med data opprettes automatisk - uten at det er nødvendig å skrive kode manuelt for hvert enkelt business case. Dette krever et infrastrukturlag som forstår datastrukturen, modellen og logikken uten ekstra instruksjoner (figur 4.1-15). Det er strukturerte data som gjør denne tilnærmingen mulig: skjemaer, tabeller, filtre og visninger kan genereres automatisk med minimal programmeringsinnsats.

De viktigste brukerkritiske grensesnittene kan fortsatt kreve manuell bearbeiding. Men i de fleste tilfeller - og dette gjelder mellom 50 og 90 prosent av arbeidsscenarioene - er det tilstrekkelig med automatisk generering av applikasjoner og beregninger uten bruk av spesialapplikasjoner for dette formålet (fig. 3.2-16), noe som reduserer utviklings- og vedlikeholdskostnadene betydelig, reduserer antall feil og fremskynder implementeringen av digitale løsninger.

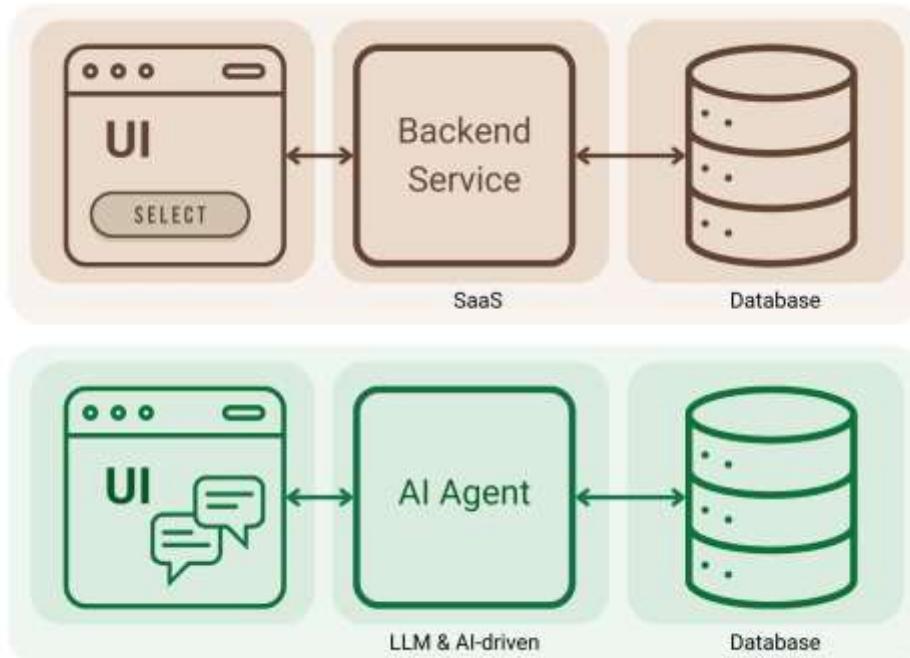
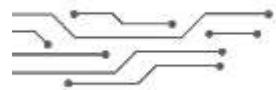


Fig. 3.2-16 Arkitekturmodeller for arbeid med data: tradisjonell applikasjonsarkitektur og AI - orientert modell med LLM.

Overgangen fra arkitekturen som bygger på individuelle applikasjoner, til intelligent styrte systemer basert på språkmodeller (LLM), er det neste steget i den digitale utviklingen. I en slik arkitektur blir strukturerte data ikke bare et lagringsobjekt, men også grunnlaget for interaksjon med AI-verktøy som er i stand til å analysere, tolke og anbefale handlinger basert på kontekst.

I de følgende kapitlene tar vi for oss eksempler fra virkeligheten på implementering av en arkitektur basert på åpne, strukturerte data og viser hvordan språkmodeller brukes til automatisk tolkning, validering og behandling av data. Disse praktiske eksemplene vil hjelpe deg å forstå hvordan den nye digitale logikken fungerer i praksis - og hvilke fordeler den gir for bedrifter som er klare for transformasjon.



KAPITTEL 3.3.

LLM OG DERES ROLLE I DATABEHANDLING OG FORRETNINGSPROSESSER

LLM chatterom: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok for å automatisere dataprosesser

Utviklingen av Large Language Models (LLM) var en naturlig forlengelse av bevegelsen mot strukturerte åpne data og Open Source-filosofien. Når data blir organisert, tilgjengelige og maskinlesbare, er neste skritt et verktøy som kan samhandle med denne informasjonen uten at man trenger å skrive kompleks kode eller besitte spesialisert teknisk kunnskap.

LLM-er er et direkte produkt av åpenhet: store åpne datasett, publikasjoner og Open Source-bevegelsen. Uten åpne vitenskapelige artikler, offentlig tilgjengelige tekstdata og en kultur for samarbeidsbasert utvikling ville det ikke eksistert noen ChatGPT eller andre LLM-er. LLM er på sitt og vis et "destillat" av menneskehets samlede digitale kunnskap, samlet og utdannet gjennom prinsippene om åpenhet.

Moderne store språkmodeller (LLM - Large Language Models) som ChatGPT ® (OpenAI), LlaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ gir brukerne muligheten til å formulere spørsmål til data på naturlig språk. Dette gjør arbeidet med informasjon tilgjengelig ikke bare for utviklere, men også for analyticere, ingeniører, planleggere, ledere og andre yrkesgrupper som tidligere var fjernt fra programmering

LLM (Large Language Model) er en kunstig intelligens som er trent opp til å forstå og generere tekst basert på store mengder data som er samlet inn fra hele Internett. Den kan analysere kontekst, svare på spørsmål, delta i dialog, skrive tekster og generere programvarekode.

Hvis visualisering, behandling eller analyse av data tidligere krevde kunnskap om et spesielt programmeringsspråk: Python, SQL, R eller Scala, samt evnen til å arbeide med biblioteker som Pandas, Polars eller DuckDB og mange andre, har situasjonen endret seg radikalt fra og med 2023. Nå kan brukeren ganske enkelt beskrive hva han eller hun ønsker å få ut - og modellen vil selv generere koden, utføre den, vise en tabell eller graf og forklare resultatet. For første gang på flere tiår har den teknologiske utviklingen ikke gått i retning av komplikasjoner, men i retning av radikal forenkling og tilgjengelighet.

Dette prinsippet - "behandle data med ord (prompter)". - markerte et nytt stadium i utviklingen av arbeidet med informasjon, og førte i praksis til at det ble mulig å skape løsninger på et enda høyere abstraksjonsnivå. På samme måte som det tidligere ikke lenger var nødvendig for brukerne å forstå det tekniske grunnlaget for Internett for å drive nettbutikker eller lage nettsteder ved hjelp av WordPress, Joomla og andre modulære systemer med åpen kildekode (av tor boken har jobbet med slike systemer siden 2005, blant annet med nettbaserte plattformer for utdanning og ingeniørarbeid). - Dette har i sin tur ført til en boom i digitalt innhold og nettbasert virksomhet - i dag kan ingeniører, analyticere og ledere automatisere arbeidsflyten uten å kunne programmeringsspråk. Dette er mulig takket være kraftige LLM-er - både gratis og med åpen

kildekode, som LLaMA, Mistral, Qwen, DeepSeek og andre - som gjør avansert teknologi tilgjengelig for et bredest mulig publikum.

Store LLM-språkmodeller: hvordan det fungerer

Store språkmodeller (ChatGPT, LLaMA, Mistral, Claude, DeepSeek, QWEN, Grok) er nevrale nettverk som er trent opp på enorme mengder tekstdata fra Internett, bøker, artikler og andre kilder. Hovedoppgaven deres er å forstå konteksten i menneskelig tale og generere meningsfulle svar.

Moderne LLM er basert på Transformer-arkitekturen som ble foreslått av Google-forskere i 2017 [85]. Den viktigste komponenten i denne arkitekturen er oppmerksomhetsmekanismen, som gjør det mulig for modellen å ta hensyn til relasjoner mellom ord uavhengig av deres posisjon i teksten.

Læringsprosessen i LLM ligner til forveksling på måten mennesker lærer språk på, bare millioner av ganger større. Modellen analyserer milliarder av eksempler på ord og uttrykk, og identifiserer mønstre i språkets struktur og i logikken i semantiske overganger. Hele teksten brytes ned i tokens - minimale semantiske enheter (ord eller deler av ord), som deretter omdannes til vektorer i et flerdimensjonalt rom (fig. 8.2-2). Disse vektorrepresentasjonene gjør det mulig for maskinen å "forstå" de skjulte relasjonene mellom begreper, i stedet for bare å behandle teksten som en sekvens av symboler.

Store språkmodeller er ikke bare verktøy for å generere tekst. De er i stand til å gjenkjenne mening, finne sammenhenger mellom begreper og arbeide med data, selv om de presenteres i ulike formater. Det viktigste er at informasjonen brytes ned i forståelige modeller og representeres som tokens som LLM-en kan arbeide med.

Den samme tilnærmingen kan brukes på byggeprosjekter. Hvis vi tenker på et prosjekt som en slags tekst, der hver bygning, hvert element eller hver konstruksjon er et symbol, kan vi begynne å behandle slik informasjon på en lignende måte. Byggeprosjekter kan sammenlignes med bøker som er organisert i kategorier, kapitler og grupper av avsnitt som består av minimale tokens - elementer i et byggeprosjekt (figur 3.3-1). Ved å oversette datamodeller til et strukturert format kan vi også oversette strukturerede data til vektorbaser (fig. 8.2-2), som er en ideell kilde for maskinlæring og teknologier som LLM.

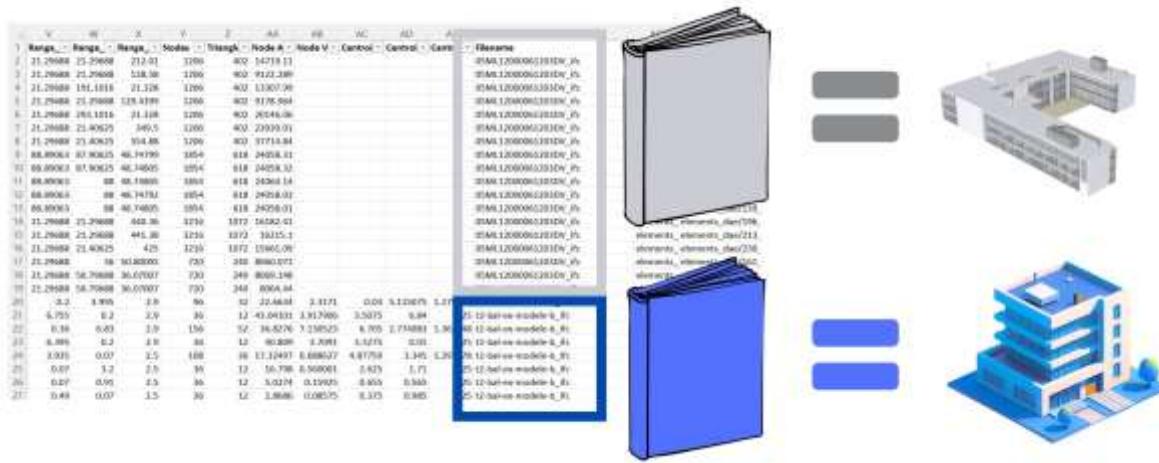


Fig. 3.3-1 Et element i et byggeprosjekt er som et symbol i en tekst: en minsteenhet som danner grupper (avsnitt) av seksjoner (kategorier) i hele prosjektet.

Hvis et byggeprosjekt er digitalisert og elementene er representert som tokens eller vektorer, blir det mulig å få tilgang til dem på naturlig språk i stedet for gjennom rigide formelle spørninger. Det er her en av de viktigste fordelene med LLM kommer inn i bildet - muligheten til å forstå betydningen av en spørring og koble den til relevante data.

Ingeniøren trenger ikke lenger å skrive SQL-forespørsler eller Python-kode for å få de nødvendige dataene - han kan ganske enkelt, med forståelse av LLM og datastrukturen, formulere oppgaven på vanlig måte: "Finn alle armerte betongkonstruksjoner med betongklasse høyere enn B30 og beregn deres totale volum". Modellen vil gjenkjenne betydningen av forespørselen, gjøre den om til en maskinlesbar form, finne dataene (gruppere og transformere) og returnere sluttresultatet.

Dokumenter, tabeller og prosjektmodeller konverteres til vektorrepresentasjoner (embedding) og lagres i databasen. Når en bruker stiller et spørsmål, konverteres også spørringen til en vektor, og systemet finner de mest relevante dataene. På denne måten kan LLM ikke bare basere seg på sin egen opplærte kunnskap, men også på faktiske bedriftsdata, selv om de allerede har dukket opp etter at modellopplæringen er avsluttet.

En av de viktigste fordelene med LLM i bygg- og anleggsbransjen er muligheten til å generere programkode. I stedet for å overlate den tekniske oppgaven til en programmerer, kan spesialister beskrive oppgaven på et naturlig språk, og modellen vil lage den nødvendige koden, som kan brukes (ved å kopiere den fra chatten) til å lage kode for prosessautomatisering. LLM-modeller gjør det mulig for spesialister uten dyp programmeringskunnskap å bidra til automatisering og forbedring av selskapets forretningsprosesser.



Figur 3.3-2 LLM-er gir brukerne muligheten til å skrive kode og få resultater uten å måtte ha programmeringsferdigheter.

Ifølge en studie utført av Wakefield Research og sponset av SAP i 2024 [36], som ble gjennomført blant 300 toppledere i selskaper med en årlig omsetning på minst 1 milliard dollar i USA: 52% av topplederne stoler på at AI analyserer data og gir anbefalinger for beslutningstaking. Ytterligere 48% bruker kunstig intelligens til å identifisere risikoer det tidligere ikke var tatt høyde for, og 47% bruker kunstig intelligens til å foreslå alternative planer. I tillegg bruker 40 prosent kunstig intelligens til utvikling av nye produkter, budsjettplanlegging og markedsundersøkelser. Studien viste også at kunstig intelligens har en positiv innvirkning på privatlivet, med 39% av respondentene som rapporterte om bedre balanse mellom arbeid og fritid, 38% som rapporterte om bedre mental helse og 31% som rapporterte om lavere stressnivå.

Men til tross for all sin kraft er LLM-er fortsatt et verktøy som det er viktig å bruke bevisst. Som all annen teknologi har de sine begrensninger. Et av de mest kjente problemene er såkalte "hallusinasjoner" - tilfeller der modellen med stor sikkerhet gir et sannsynlig, men faktamessig feilaktig svar. Det er derfor avgjørende å forstå hvordan modellen fungerer: hvilke data og datamodeller den kan tolke uten feil, hvordan den tolker spørsmål, og hvor den får informasjonen fra. Det er også verdt å huske på at LLM-modellens kunnskap er begrenset til den dataen den ble opplært, og uten en kobling til eksterne data er det ikke sikkert at modellen tar hensyn til gjeldende normer, standarder, priser eller teknologier.

Løsningen på disse problemene er å jevnlig oppdatere vektordatabaser, koble seg til relevante kilder og utvikle autonome AI-agenter - agenter som ikke bare svarer på spørsmål, men som proaktivt bruker data til opplæring, administrerer oppgaver, identifiserer risikoer, foreslår optimaliseringsalternativer og overvåker prosjekts ytelse.

Overgangen til LLM -grensesnitt i bygg- og anleggsbransjen er ikke bare en teknologisk nyhet. Det er et paradigmeskifte som fjerner barrierer mellom mennesker og data. Det handler om muligheten til å arbeide med informasjon like enkelt som vi snakker med hverandre - og likevel få nøyaktige, verifiserte og handlingsrettede resultater.

De selskapene som begynner å bruke slike verktøy tidligere enn andre, vil få et betydelig konkurransefortrinn. Dette innebærer blant annet raskere arbeid, reduserte kostnader og bedre kvalitet på

designløsningene gjennom rask tilgang til dataanalyse og muligheten til raskt å finne svar på komplekse spørsmål. Men det er også sikkerhetsspørsmål å ta hensyn til. Bruk av skybaserte LLM -tjenester kan være forbundet med risiko for datalekkasje. Derfor ser organisasjoner i økende grad etter alternative løsninger som gjør det mulig for dem å distribuere LLM-verktøy i sin egen infrastruktur - lokalt, med full beskyttelse og kontroll over informasjonen.

Bruk av lokale LLM-er for sensitive bedriftsdata

Da de første chat-LLM-ene dukket opp i 2022, markerte det et nytt stadium i utviklingen av kunstig intelligens. Men umiddelbart etter at disse modellene ble tatt i bruk i stor skala, dukket det opp et legitimt spørsmål: Hvor sikkert er det å overføre bedriftsrelaterte data og spørsmål til skyen? De fleste skybaserte språkmodeller lagret kommunikasjonshistorikk og opplastede dokumenter på serverne sine, og for bedrifter som håndterer sensitiv informasjon, var dette en alvorlig hindring for å ta i bruk kunstig intelligens.

En av de mest bærekraftige og logiske løsningene på dette problemet har vært å distribuere Open Source LLM lokalt, innenfor bedriftens IT-infrastruktur. I motsetning til skytjenester fungerer lokale modeller uten Internett-tilkobling, overfører ikke data til eksterne servere og gir bedriftene full kontroll over informasjonen.

Den beste åpne modellen [Open Source LLM] kan for øyeblikket sammenlignes i ytelse med lukkede modeller [som ChatGPT, Claude], men med et etterslep på omtrent ett år [77].

- Ben Cottier, ledende forsker ved Epoch AI, en ideell forskningsorganisasjon, 2024

Store teknologiselskaper har begynt å gjøre LLM-ene sine tilgjengelige for lokal bruk. Metas LLaMA-serie med åpen kildekode og det raskt voksende DeepSeek -prosjektet fra Kina var eksempler på overgangen til åpen arkitektur. I tillegg har Mistral og Falcon også lansert kraftige modeller som ikke er bundet av proprietære plattformer. Disse initiativene har ikke bare satt fart i utviklingen av global AI, men har også gitt personvernbevisste selskaper reelle alternativer når det gjelder uavhengighet, fleksibilitet og overholdelse av sikkerhetskrav.

I et bedriftsmiljø, spesielt i byggebransjen, er databeskyttelse ikke bare et spørsmål om bekvemmelighet, men også om overholdelse av regelverket. Arbeid med anbudsdocumenter, estimater, tegninger og konfidensiell korrespondanse krever streng kontroll. Og det er her lokal LLM gir den nødvendige forsikringen om at dataene forblir innenfor bedriftens grenser.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
Data Control	Data is transmitted to third parties	Data remains within the company's network
License	Proprietary, paid	Open-source (Apache 2.0, MIT)
Infrastructure	Requires internet	Operates in an isolated environment
Customization	Limited	Full adaptation to company needs
Cost	Pay-per-token/request	One-time hardware investment + maintenance costs
Scalability	Easily scalable with cloud resources	Scaling requires additional local hardware
Security & Compliance	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
Performance & Latency	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
Integration	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
Updates & Maintenance	Automatically updated by provider	Requires manual updates and model retraining
Energy Consumption	Energy cost is covered by provider	High power consumption for inference and training
Offline Availability	Not available without an internet connection	Works completely offline
Inference Cost	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Fig. 3.3-3 Lokale modeller gir full kontroll og sikkerhet, mens skybaserte løsninger tilbyr enkel integrering og automatiske oppdateringer.

De viktigste fordelene med lokal Open Source LLM:

- Fullstendig kontroll over data. All informasjon forblir i selskapet, noe som eliminerer uautorisert tilgang og datalekkasje.
- Frakoblet drift. Ingen avhengighet av internettforbindelse, noe som er spesielt viktig for isolerte IT-infrastrukturer. Dette sikrer også uavbrutt drift i tilfelle sanksjoner eller blokkerte skytjenester.
- Fleksible bruksområder. Modellen kan brukes til tekstgenerering, dataanalyse, skriving av programkode, designstøtte og styring av forretningsprosesser.
- Tilpasning til bedriftens mål. LLM kan læres opp på interne dokumenter, noe som gjør det mulig å ta hensyn til bedriftens spesifikke arbeid og bransjetrekk. Den lokale LLM-en kan kobles til CRM, ERP eller BI-plattformer, slik at du kan automatisere analysen av kundeforespørslar, oppretting av

rapporter eller til og med trendprognosenter.

Å distribuere DeepSeeks gratis og åpen kildekode-modell -R1-7B på en server, for tilgang for et helt team av brukere, til en kostnad på 1000 dollar per måned, kan potensielt koste mindre enn årlige avgifter for API-er i skyen, som ChatGPT eller Claude, og gjør det mulig for selskaper å ta full kontroll over dataene sine, eliminere overføring til Internett og bidra til å overholde lovpålagte krav som for eksempel GDPR

I andre bransjer er lokale LLM-er allerede i ferd med å endre sin tilnærming til automatisering. I supportjenester svarer de på hyppige kundehenvendelser, noe som reduserer arbeidsmengden til operatørene. I HR-avdelinger analyserer de CV-er og velger ut relevante kandidater. Innen e-handel genererer de persontilpassede tilbud uten å avsløre brukerdata.

En lignende effekt forventes i bygg- og anleggssektoren. Takket være integreringen av LLM med prosjektdata og standarder er det mulig å fremskynde utarbeidelsen av dokumentasjon, automatisere utarbeidelsen av estimer og prediktive kostnadsanalyser. Bruk av LLM i forbindelse med strukturerte tabeller og datarammer er i ferd med å bli et spesielt lovende område.

Full kontroll over AI i selskapet og hvordan du distribuerer din egen LLM

Moderne verktøy gjør det mulig for bedrifter å distribuere en stor språkmodell (LLM) lokalt på bare noen få timer. Dette gir full kontroll over data og infrastruktur, eliminerer avhengigheten av eksterne skytjenester og minimerer risikoen for informasjonslekkasje. Denne løsningen er spesielt relevant for organisasjoner som arbeider med sensitiv prosjektdokumentasjon eller kommersielt sensitive data.

Avhengig av oppgaver og ressurser finnes det ulike distribusjonsscenarioer, fra ferdige løsninger til mer fleksible og skalerbare arkitekturen. Et av de enkleste verktøyene er Ollama, som lar deg kjøre språkmodeller bokstavelig talt med ett klikk, uten behov for dyp teknisk kunnskap. En rask start med Ollama:

1. Last ned distribusjonen for operativsystemet ditt (Windows / Linux / macOS) fra det offisielle nettstedet: ollama.com
2. Installer modellen via kommandolinjen. For eksempel for *Mistral-modellen*:

```
ollama run mistral
```

3. Når du har kjørt modellen, er den klar til bruk - du kan sende tekstspørrengter via terminalen eller integrere den i andre verktøy. Kjør modellen og utfør en spørring:

```
ollama kjørte mistral "Hvordan lage en beregning med alle ressursene for arbeidet med å installere en 100 mm bred gipsplatevegg?"
```

For de som foretrekker å jobbe i et velkjent visuelt miljø, finnes LM Studio et gratis program med et grensesnitt som minner om ChatGPT:

- Installer LM Studio ved å laste ned distribusjonssettet fra det offisielle nettstedet - lmstudio.ai
- Velg en modell (f.eks. Falcon eller GPT-Neo-X) via den innebygde katalogen og last den ned
- Arbeid med modellen gjennom et intuitivt grensesnitt som minner om ChatGPT, men som er fullstendig lokalisert

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Figur 3.3-4 Sammenligning av populære lokale LLM-modeller med åpen kildekode -modeller.

Valg av modell avhenger av kravene til hastighet, nøyaktighet og tilgjengelige maskinvarekapasiteter (figur 3.3-4). Små modeller som Mistral 7B og Baichuan 7B egner seg for lette oppgaver og mobile enheter, mens kraftige modeller som DeepSeek -V3 krever betydelige beregningsressurser, men tilbyr høy ytelse og støtte for flere språk. I de kommende årene vil LLM-markedet vokse raskt - vi vil se stadig flere lette og spesialiserte modeller. I stedet for generelle LLM-er som dekker alt menneskelig innhold, vil det dukke opp modeller som er trent opp på smal domeneekspertise. Vi kan for eksempel forvente at det vil dukke opp modeller som kun er utviklet for å håndtere ingenørberegninger, konstruksjonsoverslag eller data i CAD-format. Slike spesialiserte modeller vil være raskere, mer nøyaktige og tryggere å bruke - særlig i profesjonelle miljøer der det er viktig med høy pålitelighet og faglig dybde.

Når den lokale LLM er lansert, kan den tilpasses til bedriftens spesifikke oppgaver. Til dette formålet brukes finjusteringsteknikken, der modellen trenes opp ytterligere på interne dokumenter, tekniske instruksjoner, kontraktsmaler eller prosjektdokumentasjon.

RAG: Intelligent LLM -assistenter med tilgang til bedriftsdata

Neste trinn i utviklingen av LLM-anvendelse i næringslivet er integrering av modeller med faktiske bedriftsdata i sanntid. Denne tilnærmingen kalles RAG (Retrieval-Augmented Generation) - gjenfinning-forsterket generering. I denne arkitekturen blir språkmodellen ikke bare et dialoggrensesnitt, men en fullverdig intelligent assistent som kan navigere i dokumenter, tegninger og databaser og gi nøyaktige, kontekstuelle svar.

Den største fordelen med RAG er muligheten til å bruke bedriftsinterne data uten å måtte trenere modellen på forhånd, samtidig som man opprettholder høy nøyaktighet og fleksibilitet i informasjonshåndteringen.

RAG -teknologien kombinerer to hovedkomponenter:

- **Henting:** Modellen kobler seg til datalagre - dokumenter, tabeller, PDF-filer - og henter relevant informasjon på forespørsel fra brukeren.
- **Augmented Generation:** Basert på de ekstraherte dataene genererer modellen et nøyaktig, informert svar som tar hensyn til konteksten og spesifisiteten til spørringen.

For å kjøre LLM med RAG-støtte, er det noen få trinn som må følges:

- **Databearbeiding:** samle inn nødvendige dokumenter, tegninger, spesifikasjoner og tabeller. De kan være i forskjellige formater og strukturer, fra PDF til Excel.
- **Indeksing og vektorisering:** Ved hjelp av verktøy som LlamaIndex eller LangChain konverteres data til vektorrepresentasjoner som gjør det mulig å finne semantiske koblinger mellom tekstrragmenter (mer om vektordatabaser og om å oversette store matriser til vektorrepresentasjon, inkludert CAD -prosjekter, i del 8).
- **Spør assistenten:** Når dataene er lastet opp, kan du stille spørsmål til modellen, og den vil søke etter svar innenfor bedriftens rammeverk i stedet for i generell kunnskap hentet fra Internett.

Anta at et selskap har en mappe constructionsite_docs, der kontrakter, instruksjoner, estimater og tabeller er lagret. Ved hjelp av et Python-skript (fig. 3.3-5) kan vi skanne denne mappen og lage en vektorindeksing: Hvert dokument blir konvertert til et sett med vektorer som gjenspeiler det semantiske innholdet i teksten. Dette gjør dokumentene til et slags "betydningskart" som modellen effektivt kan navigere i og finne forbindelser mellom termer og fraser.

Modellen "husker" for eksempel at ordene "retur" og "reklamasjon" ofte finnes i den delen av kontrakten som gjelder forsendelse av materialer til byggeplassen. Hvis det da stilles et spørsmål - for eksempel "Hva er vår returperiode?" (Fig. 3.3-5 - linje 11 i koden) - vil LLM analysere interne dokumenter og finne den nøyaktige informasjonen, og opprette som en intelligent assistent som kan lese og forstå innholdet i alle

bedriftens filer.



```

1 from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3 # Load documents from the folder
4 documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6 # Creating a vector index for semantic search
7 index = VectorStoreIndex.from_documents(documents)
8
9 # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)

```

Fig. 3.3-5 LM leser en mappe med filer - på samme måte som en person åpner den og søker etter det ønskede dokumentet

Koden kan kjøres på alle datamaskiner med Python installert. Vi skal snakke mer om bruk av Python og IDE-er for å kjøre koden i neste kapittel.

Lokal distribusjon av LLM er ikke bare en trend, men en strategisk løsning for bedrifter som verdsetter sikkerhet og fleksibilitet. Å ta i bruk LLM, enten det er på lokale bedriftsdatamaskiner eller ved hjelp av nettbaserte løsninger, er imidlertid bare første skritt. For å kunne bruke LLM-funksjonene på virkelige oppgaver må bedriftene bruke verktøy som ikke bare gjør det mulig å motta chattesvar, men som også lagrer logikken som skapes i form av kode som kan kjøres utenfor LLM-konteksten. Dette er viktig for skalering av løsninger - med riktig organiserte prosesser kan AI-utviklingen brukes på flere prosjekter samtidig eller til og med på tvers av hele selskapet.

I denne sammenhengen spiller valget av et egnet utviklingsmiljø (IDE) en viktig rolle. Moderne programmeringsverktøy på gjør det ikke bare mulig å utvikle LLM-baserte løsninger, men også å integrere dem i eksisterende forretningsprosesser og gjøre dem om til automatiserte ETL -Pipeline



KAPITTEL 3.4.

IDE MED LLM-STØTTE OG FREMTIDIGE PROGRAMMERINGSENDRINGER

Valg av IDE: fra LLM eksperimenter til forretningsløsninger

Når du skal dykke ned i en verden av automatisering, dataanalyse og kunstig intelligens - og spesielt når du arbeider med store språkmodeller (LLM-en) - er det avgjørende å velge riktig integrert utviklingsmiljø (IDE). IDE-en blir ditt viktigste arbeidsverktøy: stedet der koden som genereres av LLM-en, kjøres, både på en lokal datamaskin og i bedriftens nettverk. Valget av IDE er ikke bare avgjørende for hvor enkelt arbeidet ditt blir, men også for hvor raskt du kan gå fra eksperimentelle LLM-spørninger til fullverdige løsninger som er integrert i reelle forretningsprosesser.

IDE (Integrated Development Environment) er en allsidig byggstein på datamaskinen din for prosessautomatisering og databehandling. I stedet for å holde en sag, hammer, drill og andre verktøy hver for seg, har du en enhet som kan gjøre alt - kutte, feste, bore og til og med sjekke kvaliteten på materialene. IDE for programmerere er et enkelt sted der du kan skrive kode (i analogi med konstruksjon - lage tegninger), teste arbeidet (montering av bygningsmodeller), finne feil (som å sjekke styrken til strukturer i konstruksjonen) og kjøre det ferdige prosjektet (gangkjøring av huset).

En oversikt over populære IDE-er:

- **PyCharm®** (JetBrains) er en kraftig og profesjonell IDE for Python. Den egner seg godt til seriøse prosjekter på grunn av det store antallet innebygde funksjoner. Grunnleggende støtte for interaktive Jupyter-filer (IPYNB) er imidlertid bare tilgjengelig i betalingsversjonen, og grensesnittet kan virke overveldende for nybegynnere.

En fil med filtypen IPYNB (Interactive Python Notebook) er et format for interaktive Jupyter®-notatbøker (figur 3.4-1) der kode, visualiseringer og forklaringer er samlet i ett enkelt dokument. Dette formatet er ideelt for å lage rapporter, analyser og opplæringsscenarioer.

- **VS Code®** (Microsoft) er et raskt, fleksibelt og tilpasningsdyktig verktøy med gratis støtte for IPYNB og mange plugin-moduler. Passer for både nybegynnere og profesjonelle. Tillater integrering av GitHub Copilot og språkmodelltillegg, noe som gjør det til et godt valg for AI og datavitenskapelige prosjekter.
- **Jupyter Notebook** - Et klassisk og populært valg for eksperimentering og læring. Her kan du skrive kode, legge til forklaringer og visualisere resultater i ett og samme grensesnitt (fig. 3.4-1). Ideelt for rask testing av hypoteser, arbeid med LLM og oppretting av reproducerbare trinn for analyse av ville data. For å administrere avhengigheter og biblioteker anbefaler vi å bruke Anaconda Navigator, et visuelt grensesnitt for å administrere Python -miljøet.

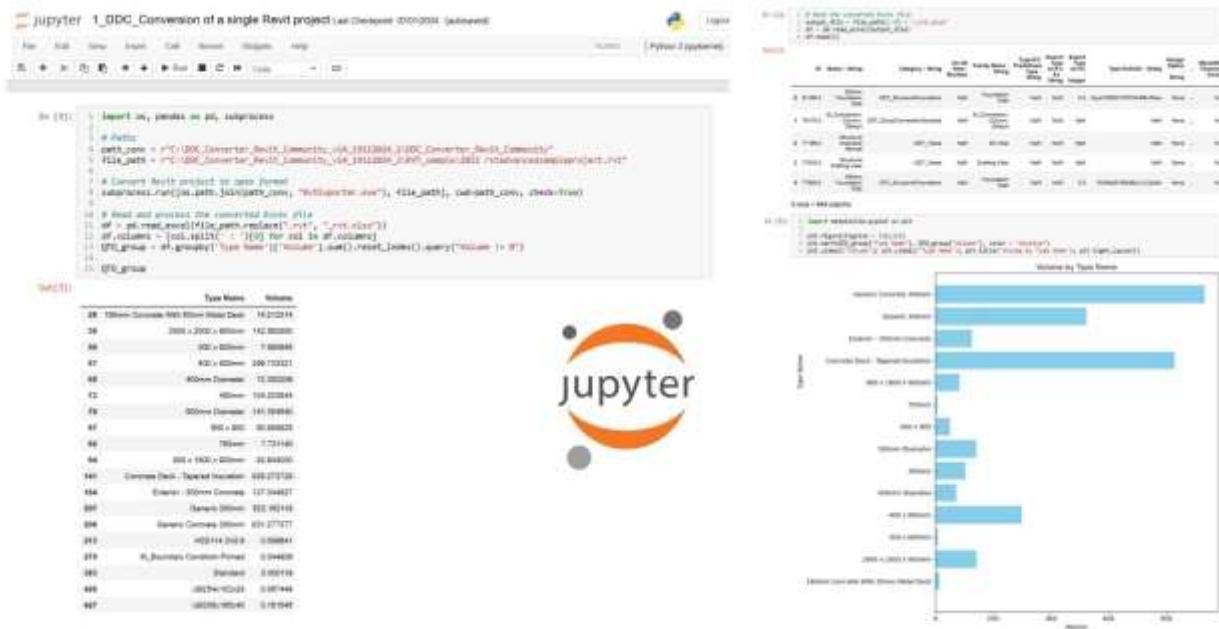


Fig. 3.4-1 Jupyter Notebook et av de mest praktiske og populære verktøyene for å lage Pipeline prosesser.

- **Google Collab™** (og Kaggle-plattformen (figur 9.2-5)) er et skybasert alternativ til Jupyter som gir gratis tilgang til GPU/TPU. Det er en flott løsning for å komme i gang - ingen lokal programvareinstallasjon og muligheten til å jobbe direkte fra en nettleser. Den støtter integrasjon med Google Drive og nylig med Gemini (Googles LLM).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Figur 3.4-2 IDE Comparison: Jupyter Notebook et av de mest praktiske og enkle verktøyene for å opprette Pipeline prosesser.

Valget av IDE avhenger av arbeidsoppgavene dine. Hvis du raskt vil begynne å jobbe med AI, kan du prøve Jupyter Notebook eller Google Collab. For seriøse prosjekter er det bedre å bruke PyCharm eller VS Code. Det viktigste er å komme i gang. Med moderne verktøy kan du raskt gjøre eksperimentene dine om til fungerende løsninger.

Alle IDE-ene som er beskrevet ovenfor, gjør det mulig å opprette databehandlingsrørledninger - det vil si kjeder av kodeblokkmoduler (som kan genereres av LLM), der hver av dem for eksempel er ansvarlig for et annet trinn:

- analytiske scenarier,
- kjeder for informasjonsutvinning fra dokumenter,
- automatiske svar basert på RAG,
- generering av rapporter og visualiseringer.

Takket være den modulære strukturen kan hvert trinn representeres som en egen blokk: innlasting av data → filtrering → analyse → visualisering → eksport av resultater. Disse blokkene kan gjenbrukes, - tilpasses og settes sammen til nye kjeder som en konstruktør, bare for data.

For ingeniører, ledere og analyticere åpner dette for muligheten til å dokumentere beslutningslogikken i form av kode som kan genereres med LLM. Denne tilnærmingen bidrar til å fremskynde rutineoppgaver, automatisere typiske operasjoner og skape repeterbare prosesser der hvert trinn er tydelig dokumentert og gjennomsiktig for alle teammedlemmer.

De automatiserte ETL-verktøyene Pipelines (fig. 7.2-3), Apache Airflow (fig. 7.4-4), Apache NiFi (fig. 7.4-5) og n8n (fig. 7.4-6) for å bygge logikkblokker for prosessautomatisering vil bli nærmere omtalt i bokens del 7 og del 8.

IDE med LLM-støtte og fremtidige programmeringsendringer

Integreringen av kunstig intelligens i utviklingsprosesser er i ferd med å endre programmeringslandskapet. Moderne miljøer er ikke lenger bare tekstdredigeringsprogrammer med syntaksutheving - de er i ferd med å bli intelligente assistenter som er i stand til å forstå prosjektlogikken, fullføre koden og til og med forklare hvordan et bestemt kodefragment fungerer. Det dukker opp produkter på markedet som bruker kunstig intelligens til å flytte grensene for konvensjonell utvikling:

- **GitHub Copilot** (integreres med VS Code, PyCharm): AI-assistent som genererer kode basert på kommentarer eller delvise beskrivelser, og gjør tekstlige hint om til ferdige løsninger.
- **Cursor** (en gaffel av VS Code med AI-kernel): gjør det ikke bare mulig å skrive ferdig kode, men også å stille spørsmål til prosjektet, se etter avhengigheter og lære av kodebasen.
- **JetBrains AI Assistant**: en plugin for JetBrains IDE (inkludert PyCharm) med funksjon for å forklare kompleks kode, optimalisering og oppretting av tester.
- **Amazon CodeWhisperer**: en analog til Copilot med fokus på sikkerhet og støtte for Amazons AWS-tjenester.

Programmering vil gjennomgå en dramatisk endring i årene som kommer. Hovedfokuset vil skifte fra rutinemessig kodeskriving til design av modeller og dataarkitektur - utviklere vil bli mer involvert i systemdesign, mens AI vil overta maloppgaver: kodegenerering, tester, dokumentasjon og grunnleggende funksjoner. Fremtidens programmering er et samarbeid mellom mennesker og kunstig intelligens, der maskinene tar over de tekniske rutinene og menneskene fokuserer på kreativiteten.

Programmering på naturlig språk vil bli en del av hverdagen. IDE-personalisering vil nå et nytt nivå - utviklingsmiljøene vil lære seg å tilpasse seg brukerens arbeidsstil og bedrifter ved å forutse mønstre, tilby

kontekstuelle løsninger og lære av tidligere prosjekter

Dette betyr ikke at utviklerens rolle forsvinner, men det vil endre den radikalt: fra å skrive kode til kunnskaps-, kvalitets- og prosesshåndtering. Denne utviklingen vil også påvirke business intelligence, der utarbeidelsen av rapporter, visualiseringer og beslutningsstøtteapplikasjoner i økende grad vil skje gjennom generering av kode og logikk ved hjelp av AI og LLM, chat og agentgrensesnitt.

Når en bedrift har satt opp LLM-chatter og valgt et egnet utviklingsmiljø, er det neste viktige steget å organisere dataene. Denne prosessen innebærer å hente ut informasjon fra ulike kilder, rense den, omforme den til en strukturert form og integrere den i bedriftens systemer.

I en moderne datasentrisk tilnærming til datahåndtering er det et sentralt mål å samle data i én universell form som er kompatibel med et stort antall verktøy og applikasjoner. For å håndtere strukturéringsprosesser og strukturerte data er det behov for spesialiserte biblioteker. Et av de kraftigste, mest fleksible og populære er Pandas-biblioteket for Python. Det gjør det enkelt å behandle tabelldata: filtrere, gruppere, rense, legge til, aggregere og rapportere.

Python Pandas: et uunnværlig verktøy for arbeid med data

Pandas har en spesiell plass i verdenen av dataanalyse og automatisering. Det er et av de mest populære og mest brukte bibliotekene i programmeringsspråket Python [86], utviklet for å arbeide med strukturerte data.

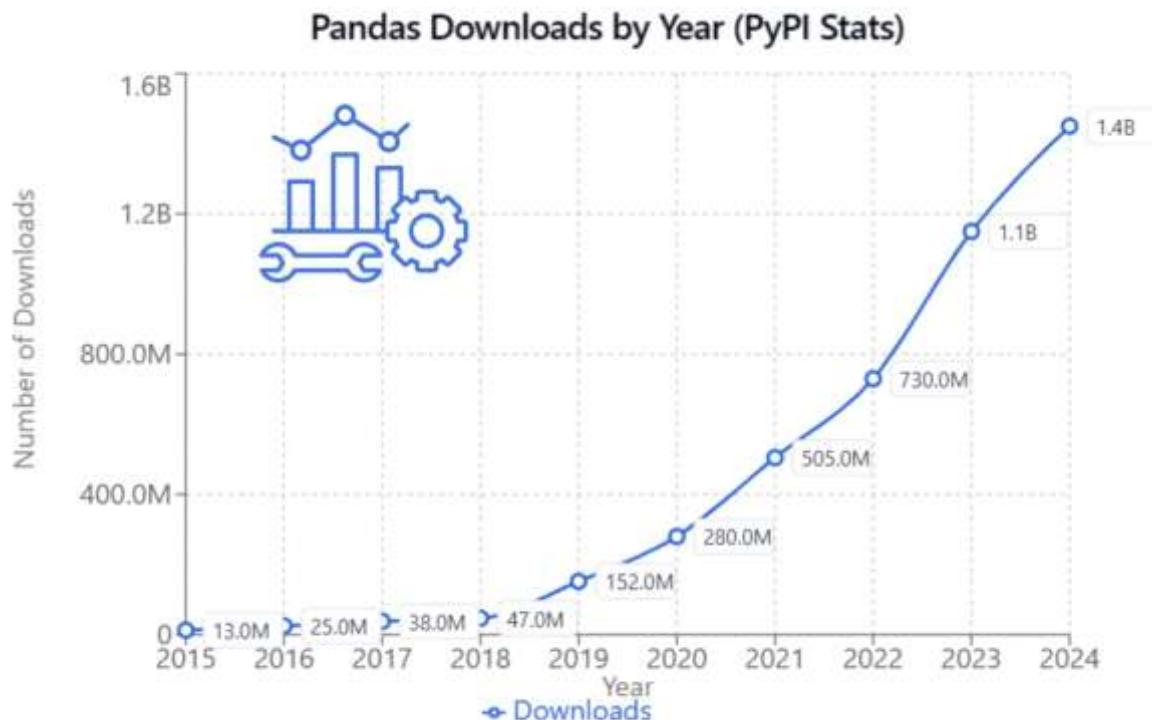
Et bibliotek er som et sett med ferdige verktøy: funksjoner, moduler, klasser. På samme måte som man på en byggeplass ikke trenger å finne opp en hammer eller et vater hver gang, gjør biblioteker det mulig å løse problemer raskt uten å måtte finne opp grunnleggende funksjoner og løsninger på nytt.

Pandas er et Python-bibliotek med åpen kildekode , som tilbyr høy ytelse og intuitive datastrukturer, spesielt DataFrame, et universelt format for arbeid med tabeller. Pandas er en sveitsisk kniv for analytikere, ingeniører og utviklere som arbeider med data.

Python er et programmeringsspråk på høyt nivå med en enkel syntaks som brukes aktivt innen analyse, automatisering, maskinlæring og webutvikling. Språket er populært på grunn av sin lesbare kode, plattformovergripende natur og et rikt økosystem av biblioteker. Til dags dato er det opprettet mer enn 137 000 åpen kildekode-pakker for Python [87], og dette antallet fortsetter å vokse nesten daglig. Hvert av disse bibliotekene er et slags lager med ferdige funksjoner: fra enkle matematiske operasjoner til komplekse verktøy for bildebehandling, stordataanalyse, nevrale nettverk og integrering med eksterne tjenester.

Forestill deg med andre ord at du har fri og åpen tilgang til hundretusener av ferdige programvareløsninger - biblioteker og verktøy som du kan bygge direkte inn i forretningsprosessene dine. Det er som en enorm katalog med applikasjoner for automatisering, analyse, visualisering, integrering og mye mer - og alt er tilgjengelig umiddelbart etter at du har installert Python.

Pandas er en av de mest populære pakkene i Python-økosystemet. I 2022 nådde gjennomsnittlig antall nedlastingar av Pandas-biblioteket 4 millioner per dag (figur 3.4-3), mens tidlig i 2025 hadde dette tallet økt til 12 millioner nedlastingar per dag, noe som gjenspeiler den økende populariteten og utbredte bruken av Pandas i dataanalyse og LLM chats [86]



Figur 3.4-3 Pandas er et av de mest nedlastede bibliotekene. I 2024 oversteg det årlige antallet nedlastingar 1,4 milliarder.

Spørrespråket i Pandas-biblioteket har samme funksjonalitet som spørrespråket SQL, som vi omtalte i kapittelet "Relasjonsdatabaser og spørrespråket SQL".

I en verden av analyser og strukturert datahåndtering skiller Pandas seg ut med sin enkelhet, hastighet og kraft, og gir brukerne et bredt spekter av verktøy for å analysere og behandle informasjon på en effektiv måte.

Begge verktøyene - SQL og Pandas - gir kraftige muligheter for datamanipulering, spesielt sammenlignet med tradisjonell Excel. De støtter operasjoner som utvelgelse og filtrering (figur 3.4-4), men den eneste forskjellen er at SQL er optimalisert for arbeid med relasjonsdatabaser, mens Pandas behandler data i RAM, noe som gjør at det kan kjøres på hvilken som helst datamaskin, uten at det er nødvendig å opprette databaser og implementere en egen infrastruktur.



Fig. 3.4-4 Pandas, i motsetning til SQL, har fleksibilitet til å arbeide med en rekke dataformater, ikke bare databaser.

Pandas er ofte foretrukket for vitenskapelig forskning, prosessautomatisering, oppretting av pipelines (inkludert ETL) og datamanipulering i Python, mens SQL er en standard for databasehåndtering og ofte brukes i bedriftsmiljøer til å håndtere store datamengder.

Pandas-biblioteket i programmeringsspråket Python gjør det mulig å utføre ikke bare grunnleggende operasjoner som å lese og skrive tabeller, men også mer komplekse oppgaver, som å slå sammen data, gruppere data og utføre komplekse analytiske beregninger.

I dag brukes Pandas -biblioteket ikke bare i akademisk forskning og forretningsanalyse, men også i forbindelse med LLM -modeller. For eksempel la Meta®-divisjonen (Facebook™), da den publiserte den nye open source-modellen LLaMA 3.1 i 2024, spesiell vekt på å arbeide med strukturerte data, og en av de viktigste og første sakene i utgivelsen var nettopp behandling av strukturerte datarammer (fig. 3.4-5) i CSV-format og integrering med Pandas-biblioteket direkte i chat.

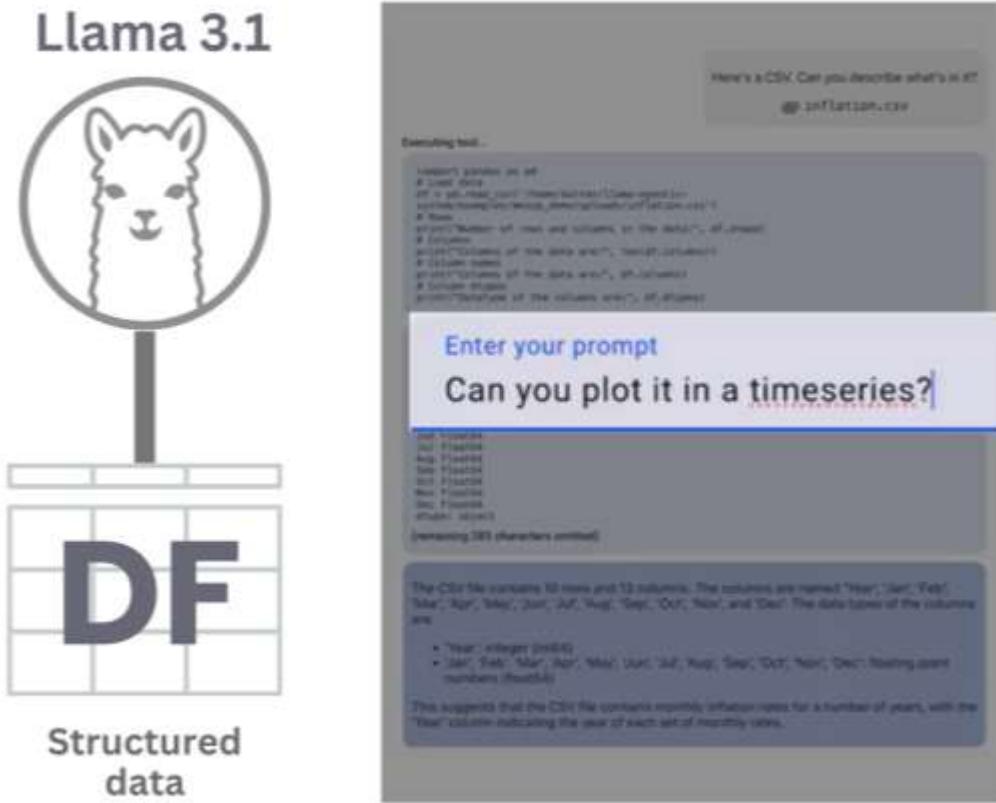


Fig. 3.4-5 Et av Meta-teamets første og viktigste case som ble presentert i LlaMa 3.1 i 2024, var å bygge applikasjoner ved hjelp av Pandas

Pandas er et viktig verktøy for millioner av dataforskere som behandler og forbereder data for generativ AI. Å akselerere Pandas uten å endre koden vil være et stort skritt fremover. Dataforskere vil kunne behandle data på minutter i stedet for timer og få mange ganger mer data til å trenne generative AI-modeller [88].

- Jensen Huang, grunnlegger og administrerende direktør i NVIDIA

Ved hjelp av Pandas er det mulig å håndtere og analysere datasett som er langt større enn det Excel kan. Mens Excel vanligvis kan håndtere opptil 1 million rader med data, kan Pandas enkelt håndtere datasett (figur 9.1-2, figur 9.1-10) som inneholder flere titalls millioner rader [89]. Dette gjør det mulig for brukerne å utføre sofistikerte dataanalyser og visualiseringer på store datasett, noe som gir dyp innsikt og gjør det enklere å ta datadrevne beslutninger. I tillegg har Pandas sterkt støtte fra fellesskapet [90]: hundrevis av millioner av utviklere og analytikere over hele verden (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) bruker det online eller offline hver dag, noe som gir et stort antall ferdige løsninger for ethvert forretningsproblem.

Kjernen i de fleste analytiske prosesser i Python er en strukturert form for data som kalles DataFrame, som leveres av Pandas-biblioteket. Det er et kraftig og fleksibelt verktøy for organisering, analyse og visualisering

av tabelldata.

DataFrame: universelt tabellformat for data

DataFrame er den sentrale strukturen i Pandas-biblioteket, som er en todimensjonal tabell (figur 3.4-6) der radene tilsvarer individuelle objekter eller poster, og kolonnene tilsvarer deres egenskaper, parametere eller kategorier. Denne strukturen ligner visuelt på Excel-regneark, men er langt bedre når det gjelder fleksibilitet, skalerbarhet og funksjonalitet.

En **DataFrame** er en måte å representere og behandle tabelldata som er lagret i datamaskinens RAM-minne.

DataFrame er en måte å representere og behandle tabelldata som er lagret i datamaskinens RAM. I en tabell kan rader for eksempel gjenspeile elementer i et byggeprosjekt, og kolonner - deres egenskaper: kategorier, dimensjoner, koordinater, kostnader, vilkår og så videre. En slik tabell kan dessuten inneholde både informasjon om ett prosjekt (fig. 4.1-13) og data om millioner av objekter fra tusenvis av ulike prosjekter (fig. 9.1-10). Takket være vektoriserte Pandas-operasjoner er det enkelt å filtrere, gruppere og aggregere slike informasjonsmengder i høy hastighet.

The diagram illustrates a DataFrame as a 2D table with columns and rows. The columns are labeled with names: ID, Name, Category, Family Name, Height, BoundingBoxMin_X, BoundingBoxMin_Y, BoundingBoxMin_Z, and Level. The rows are indexed from 431144 to 497540. Annotations highlight specific features:

- Index label:** Points to the first column, labeled "Index label".
- Index axis = 0:** Points to the first row, labeled "Index axis = 0".
- Column names:** Points to the column headers.
- Missing value:** Points to a cell containing a red box, labeled "Missing value".
- Data:** Points to a green box spanning multiple cells, labeled "Data".
- Level:** Points to the last column, labeled "Level".

ID	Name	Category	Family Name	Height	BoundingBoxMin_X	BoundingBoxMin_Y	BoundingBoxMin_Z	Level
431144	Single-Flush	OST_Doors	Single-Flush	6.88976378	20.1503	-10.438	9.84252	Level 1
431198	Single-Flush	OST_Doors		6.88976378	13.2281	-1.1207	9.84252	Level 2
457479	Single Window	OST_Windows	Single Window		-11.434	-11.985	9.80971	Level 2
485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971	
490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.5748	-2.9565	-1E-16	Level 1
493697	Basic Wall	OST_Walls	Basic Wall		-38.15	20.1656	-4.9213	Level 1
497540	Basic Wall	OST_Walls	Basic Wall		4.5212	0.0708	9.84252	Level 1

Figur 3.4-6 Construction Project som en DataFrame er en todimensjonal tabell med elementer i rader og attributter i kolonner.

Nvidia anslår at opptil 30 prosent av alle databehandlingsressurser allerede i dag brukes til å behandle strukturerte data - dataframes - og at denne andelen fortsetter å vokse.

Databehandling er det som trolig står for en tredjedel av verdens databehandling i hver eneste bedrift. Databehandlingen og dataene i de fleste selskaper er i DataFrame-tabellformat.

- Jensen Huang, administrerende direktør i Nvidia [91]

La oss liste opp noen viktige funksjoner i DataFrame i Pandas:

- **Kolonner:** I DataFrame er data organisert i kolonner, hver med et unikt navn. Attributtkolonner kan inneholde data av ulike typer, på samme måte som kolonner i databaser eller kolonner i tabeller.
 - **Pandas Series** er en endimensjonal datastruktur i Pandas, på samme måte som en liste eller kolonne i en tabell, der hver verdi tilsvarer en annen indeks
- Pandas Series har over 400 attributter og metoder som gjør det utrolig fleksibelt å arbeide med data. Du kan bruke en av de 400 tilgjengelige funksjonene direkte på en kolonne, utføre matematiske operasjoner, filtrere data, erstatte verdier, jobbe med datoer, strenger og mye mer. I tillegg støtter Series vektoriserte operasjoner, noe som gjør behandlingen av store datasett mye raskere enn ved sykliske beregninger. Du kan for eksempel enkelt multiplisere alle verdier med et tall, erstatte manglende data eller bruke komplekse transformasjoner uten å skrive kompliserte løkker.
- **Rader:** I DataFrame kan indekseres med unike verdier. Med denne indeksen kan du raskt endre og justere dataene i bestemte rader.
 - **Indeks:** Når du oppretter en DataFrame Pandas tilordner du som standard hver rad en indeks fra 0 til N-1 (der N er antallet rader i DataFrame). Indeksen kan imidlertid endres for å inkludere spesielle betegnelser som datoer eller unike egenskaper.
 - **Indeksering** av rader i en DataFrame betyr at hver rad tildeles et unikt navn eller en etikett, som kalles DataFrame-indeksen.
 - **Datatyper:** DataFrame støtter en rekke datatyper, inkludert `int`, `float`, `bool`, `datetime64` og `object` for tekstdata. Hver DataFrame-kolonne har sin egen datatype som bestemmer hvilke operasjoner som kan utføres på innholdet.
 - **Dataoperasjoner:** DataFrame støtter et bredt spekter av operasjoner for databehandling, inkludert aggregering (`groupby`), sammenslåing (`merge` og `join`), konkatenering (`concat`), split-apply-combine og mange andre datatransformasjonsteknikker.
 - **Størrelsesmanipulering:** DataFrame lar deg legge til og fjerne kolonner og rader, noe som gjør den til en dynamisk struktur som kan endres i henhold til dataanalysebehovene dine.
 - **Visualisering av data:** Ved hjelp av innebygde visualiseringsteknikker eller grensesnitt med populære datavisualiseringsbiblioteker som Matplotlib eller Seaborn kan DataFrame enkelt konverteres til grafer og diagrammer for å presentere data grafisk.
 - **Datainput og -output:** Pandas har funksjoner for å lese, importere og eksportere data til ulike filformater som CSV, Excel, JSON, HTML og SQL, noe som potensielt kan gjøre DataFrame til et sentralt knutepunkt for datainnsamling og -distribusjon.

I motsetning til CSV og XLSX gir Pandas DataFrame større fleksibilitet og ytelse når du arbeider med data: Den kan håndtere store mengder informasjon i RAM, støtter utvidede datatyper (inkludert datoer, logiske verdier og tidsserier) og gir omfattende muligheter for filtrering, aggregering, sammenslåing og visualisering av data. Selv om CSV ikke lagrer informasjon om datatyper og -struktur, og XLSX ofte er overbelastet med formatering og har lav skalerbarhet, er DataFrame fortsatt det optimale valget for rask analyse, prosessautomatisering og integrering med AI -modeller (figur 3.4-7). I de følgende kapitlene vil hvert av disse aspektene ved data bli utforsket i detalj, og i del 8 av boken vil lignende formater som Parquet, Apache Orc, JSON, Feather, HDF5 og datavarehus bli diskutert i detalj (fig. 8.1-2).

		XLSX	CSV	Pandas DataFrame
	Storage	Tabular	Tabular	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
	Compression	Built-in	None	None (in-memory)
	Performance	Low	Medium	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Low
	Data Type Support	Limited	Very limited	Extended
	Scalability	Low	Low	Medium (memory limited)

Fig. 3.4-7 DataFrame er det optimale valget for datamanipulering med høy ytelse og avansert støtte for datatyper.

Pandas-biblioteket og DataFrame-formatet har takket være sin fleksibilitet, kraft og brukervennlighet blitt de facto-standarden for dataanalyse i Python. De er ideelle for både å lage enkle rapporter og bygge komplekse analytiske pipelines, spesielt i forbindelse med LLM-modeller.

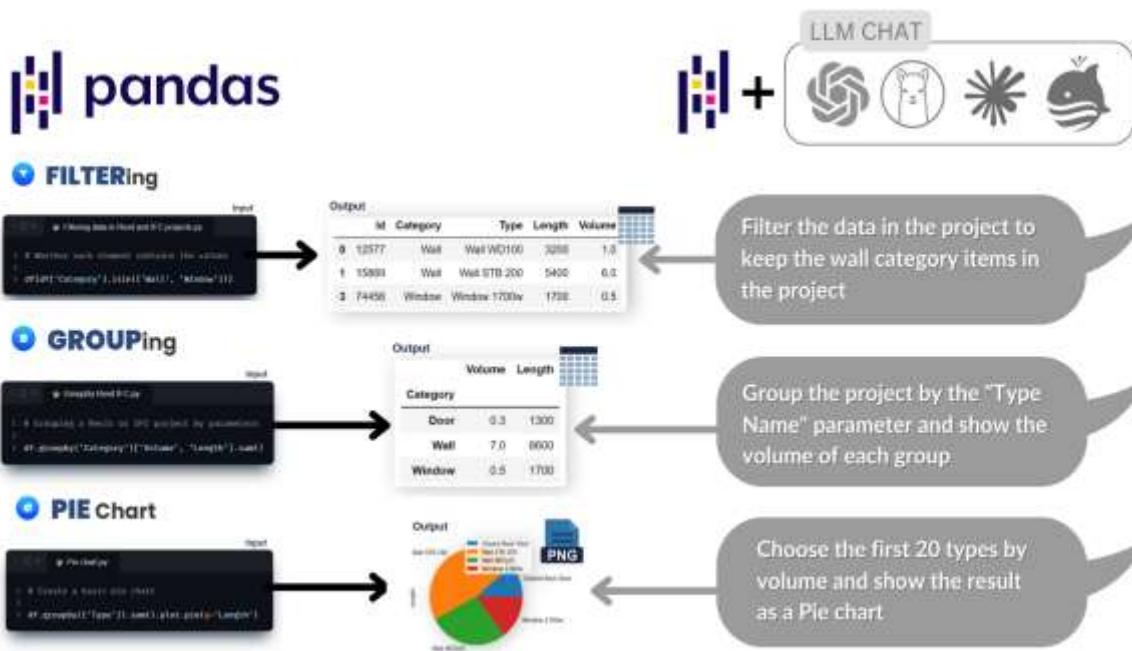


Fig. 3.4-8 LLM-er forenkler interaksjonen med Pandas: I stedet for kode er det tilstrekkelig med en tekstforespørsel.

I dag brukes Pandas aktivt i LLM-baserte chatterom som ChatGPT, LLaMa, DeepSeek, QWEN og andre. I mange tilfeller, når en modell mottar en spørring relatert til tabellbehandling, datavalidering eller analyse, genererer den kode nøyaktig ved hjelp av Pandas-biblioteket. Dette gjør DataFrame til et naturlig "språk" for å representere data i dialoger med kunstig intelligens (figur 3.4-8).

Moderne data teknologi som Pandas gjør det enklere å analysere, automatisere og integrere data i forretningsprosesser. De leverer resultater raskt, reduserer arbeidsmengden for spesialister og sikrer repeterbare operasjoner.

Neste steg: å bygge et bærekraftig datarammeverk

I denne delen så vi på de viktigste datatypene som brukes i byggebransjen, lærte om de ulike formatene for lagring av data og analyserte hvilken rolle moderne verktøy, inkludert LLM og IDE-er, spiller i behandlingen av informasjon. Vi lærte at effektiv datahåndtering er grunnlaget for informert beslutningstaking og automatisering av forretningsprosesser. Organisasjoner som er i stand til å strukturere og organisere dataene sine, får et betydelig konkurransefortrinn i databehandlings- og transformasjonsfasen.

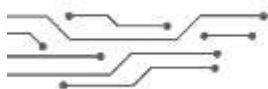
For å oppsummere denne delen er det verd å fremheve de viktigste praktiske trinnene som vil hjelpe deg med å bruke de diskuterte tilnærmingene i det daglige arbeidet:

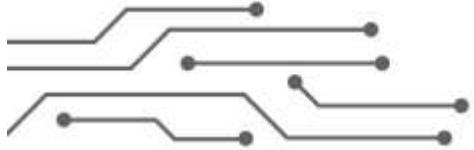
- Gjennomfør en data revisjon av prosessene dine
- Lag en oversikt over alle datatypene du bruker i prosjektene dine
- Bestem hvilke datatyper og modeller som er mest kritiske for forretningsprosessene dine

- Identifisere problemområder der informasjon ofte forblir ustrukturert, dårlig struktureret eller utilgjengelig
- Begynn å utforme en strategi for datahåndtering
 - Ta opp politiske spørsmål og standarder for håndtering av ulike datatyper
 - Analyser hvilke arbeidsflyter som kan forbedres ved å konvertere ustrukturerte data til strukturerte data
 - Utarbeide retningslinjer for datalagring og -tilgang som tar hensyn til sikkerhet og konfidensialitet
- Installere og beherske grunnleggende verktøy for arbeid med data
 - Velg en IDE som passer til oppgavene dine (installer f.eks. VS Code eller Jupyter Notebook)
 - Prøv å opprette en lokal LLM for å håndtere personopplysningene dine konfidensielt
 - Begynn å eksperimentere med Pandas-biblioteket for å behandle XLSX-tabelldata
 - Beskriv de typiske oppgavene du håndterer i regnearkverktøy eller databaser, og be LLM-en om å automatisere arbeidet ved hjelp av Pandas

Ved å følge disse trinnene kan du gradvis endre måten du arbeider med data på, og gå fra ulike, ustrukturerte datasett til et enhetlig økosystem der data blir en tilgjengelig og forståelig ressurs. Begynn i det små - lag din første DataFrame i Pandas, kjør en lokal LLM, automatiser din første rutineoppgave ved hjelp av Python (f.eks. Excel-regneark).

Den fjerde delen av boken fokuserer på datakvalitet, organisering, strukturering og modellering av data. Vi fokuserer på metoder som omdanner ulike datakilder - fra PDF-filer og tekster til bilder og CAD-modeller - til strukturerte datasett som egner seg for analyse og automatisering. Vi vil også se nærmere på hvordan datakrav formaliseres, hvordan konseptuelle og logiske modeller bygges opp i byggeprosjekter, og hvordan moderne språkmodeller (LLM-er) kan bidra i denne prosessen.





IV DEL

DATAKVALITET: ORGANISERING, STRUKTURERING, MODELLERING

Den fjerde delen fokuserer på metoder og teknologier for å omdanne uensartet informasjon til strukturerte datasett av høy kvalitet. Prosessene for å utforme og dokumentere datakrav som grunnlag for en effektiv informasjonsarkitektur i byggeprosjekter blir diskutert i detalj. Praktiske metoder for å hente ut strukturert informasjon fra ulike kilder (PDF -dokumenter, bilder, tekstmater, CAD -modeller) presenteres med eksempler på implementering. Bruken av regulære uttrykk (RegEx) og andre verktøy for automatisk validering og verifisering av data blir analysert. Prosessen med datamodellering på konseptuelt, logisk og fysisk nivå beskrives trinn for trinn, med hensyn til de spesielle forholdene i byggebransjen. Det demonstreres konkrete eksempler på bruk av språkmodeller (LLM) for å automatisere strukturering og validering av informasjon. Det foreslås effektive metoder for visualisering av analyseresultater, noe som øker tilgjengeligheten av analytisk informasjon for alle nivåer i ledelsen av byggeprosjekter

KAPITTEL 4.1.

KONVERTERING AV DATA TIL EN STRUKTURERT FORM

I den datadrevne økonomiens tidsalder er data i ferd med å bli grunnlaget for beslutningstaking snarere enn et hinder. I stedet for å hele tiden tilpasse informasjonen til nye systemer og deres formater, søker bedrifter i økende grad å skape én enkelt strukturert datamodell som fungerer som en universell sannhetskilde for alle prosesser. Moderne informasjonssystemer er ikke utformet rundt formater og grensesnitt, men rundt betydningen av data - fordi strukturen kan endres, men betydningen av informasjonen forblir den samme i mye lengre tid.

Nøkkelen til å jobbe effektivt med data er ikke å konvertere og transformere dem i det uendelige, men å organisere dem riktig fra begynnelsen av. Det handler om å skape en universell struktur som gir åpenhet, automatisering og integrering i alle faser av prosjektets livssyklus.

Den tradisjonelle tilnærmingen tvinger frem manuelle justeringer ved hver nye plattformimplementering: migrering av data, endring av attributtnavn, justering av formater. Disse trinnene forbedrer ikke selve datakvaliteten, men maskerer bare problemene og skaper en ond sirkel av endeløse transformasjoner. Resultatet er at bedriftene blir avhengige av spesifikke programvareløsninger, og den digitale transformasjonen bremses opp.

I de neste kapitlene skal vi se nærmere på hvordan man strukturerer data på riktig måte, og deretter hvordan man lager universelle modeller, minimerer plattformavhengigheten og fokuserer på det som betyr mest - data som en strategisk ressurs som bærekraftige prosesser bygges rundt.

Lære hvordan du gjør dokumenter, PDF-filer, bilder og tekster om til strukturerte formater

I byggeprosjekter finnes det aller meste av informasjonen i ustukturert form: tekniske dokumenter, arbeidsbeskrivelser, tegninger, spesifikasjoner, tidsplaner og protokoller. Mangfoldet - både når det gjelder format og innhold - gjør det vanskelig å integrere og automatisere.

Prosessens med å konvertere til strukturerte eller halvstrukturerte formater kan variere avhengig av typen inndata og ønsket behandlingsresultat.

Å transformere data fra ustukturert til strukturert form er både en kunst og en vitenskap. Denne prosessen varierer avhengig av typen inngangsdata og formålet med analysen, og tar ofte en betydelig del av arbeidet til dataingeniøren (figur 3.2-5) og analytikeren, med mål om å produsere et rent, organisert datasett.



Fig. 4.1-1 Konvertering av et ustrukturert skannet dokument til et strukturert tabellformat.

Å gjøre om dokumenter, PDF-filer, bilder og tekster til et strukturert format (figur 4.1-1) er en trinnvis prosess som omfatter følgende trinn:

- **Utdrag:** I dette trinnet lastes et kildedokument eller et bilde som inneholder ustrukturerte data. Dette kan for eksempel være et PDF -dokument, et fotografi, en tegning eller et skjema.
- **Datatransformasjon (Transform):** Deretter følger et trinn der ustrukturerte data omdannes til et strukturert format. Dette kan for eksempel innebære å gjenkjenne og tolke tekst fra bilder ved hjelp av optisk tegngjenkjenning (OCR) eller andre behandlingsmetoder.
- **Laste inn og lagre data (Load):** Det siste trinnet innebærer å lagre de bearbeide dataene i ulike formater, for eksempel CSV, XLSX, XML, JSON, for videre arbeid, der valget av format avhenger av spesifikke krav og preferanser.

Denne prosessen, kjent som ETL (Extract, Transform, Load), spiller en nøkkelrolle i automatisert databehandling og vil bli diskutert mer inngående i kapittelet "ETL og pipeline: Extract, Transform, Load". Deretter skal vi se på eksempler på hvordan dokumenter i ulike formater transformeres til strukturerte data.

Eksempel på konvertering av et PDF -dokument til en tabell

En av de vanligste oppgavene i byggeprosjekter er å behandle spesifikasjoner i PDF-format. For å demonstrere overgangen fra ustrukturerte data til et strukturert format kan vi ta for oss et praktisk eksempel: å trekke ut en tabell fra et PDF-dokument og konvertere den til CSV- eller Excel-format (fig. 4.1-2).

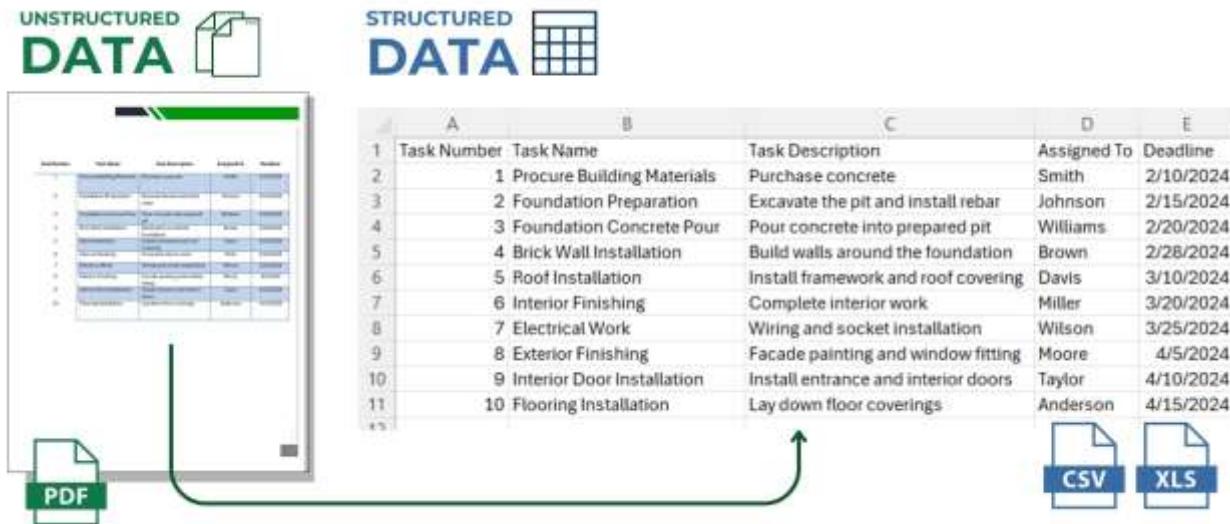


Fig. 4.1-2 I motsetning til PDF er CSV og XLSX formatene utbredt og kan enkelt integreres i ulike datahåndteringssystemer.

LLM-språkmodeller, som ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN forenkler i stor grad måten dataforskere arbeider med data på, reduserer behovet for dyp læring av programmeringsspråk og gjør det mulig å løse mange oppgaver med tekstspøringer.

I stedet for å bruke tid på å lete etter løsninger på Internett (vanligvis StackOverflow-nettstedet eller tematiske fora og chatter) eller kontakte databehandlingsspesialister, kan vi derfor bruke funksjonene til moderne online eller lokale LLM-er. Det er nok å stille et spørsmål, og modellen vil gi klar kode for å konvertere PDF -dokument til et tabellformat.

- Send følgende tekstforespørsel til en hvilken som helst LLM -modell (CHATGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN eller en hvilken som helst annen):

Vennligst skriv en kode for å trekke ut tekst fra en PDF -fil som inneholder en tabell. Koden skal ta filstien som et argument og returnere den ekstraherte tabellen som en DataFrame.

- LLM -modellsvaret vil i de fleste tilfeller være i form av kode i Python , ettersom dette språket er mye brukt til databehandling, automatisering og arbeid med ulike filformater:

```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11         Extract text from a PDF file.
12         Args: pdf_path (str): Path to the PDF file.
13         Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

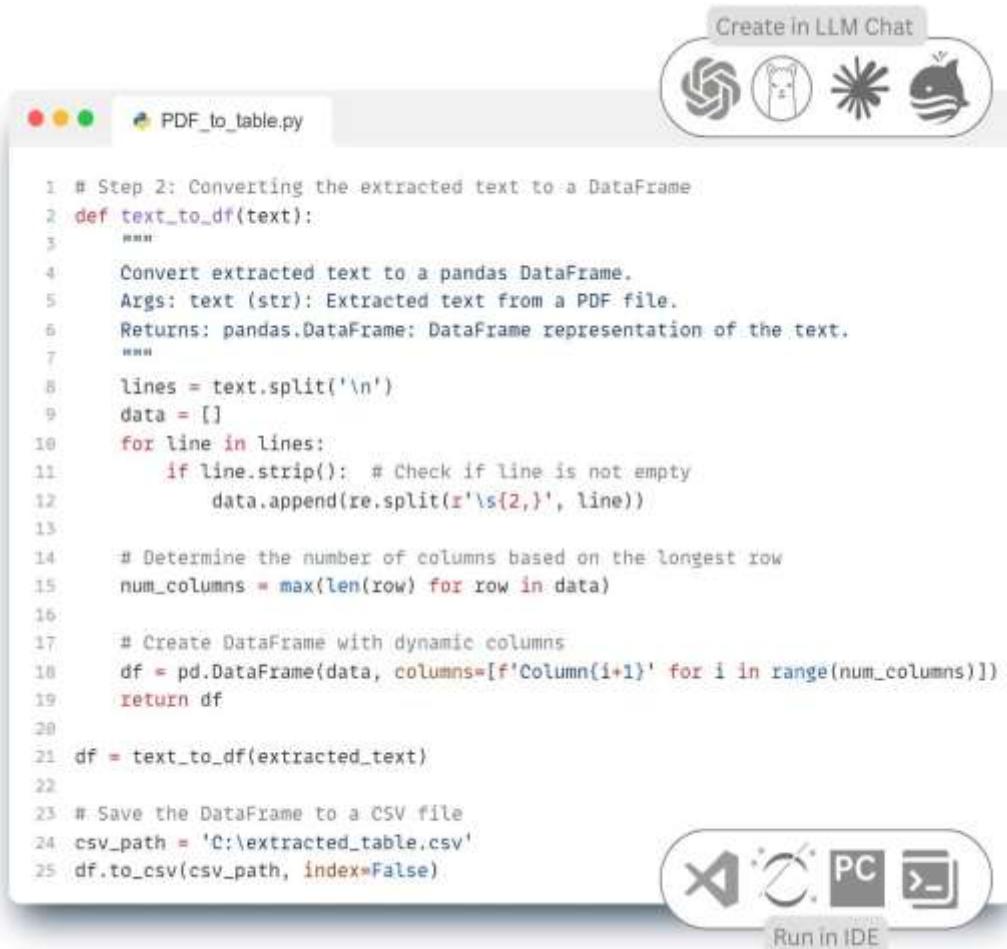
Fig. 4.1-3 LLM-svaret i form av Python-koden og tilhørende biblioteker og pakker (Pandas, Fitz) trekker ut tekst fra en PDF -fil.

Denne koden (figur 4.1-3) kan kjøres offline i en av de populære IDE-ene vi nevnte ovenfor: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære nettbaserte verktøy: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- I "Convert"-trinnet bruker vi det populære Pandas-biblioteket (som vi diskuterte i detalj i kapittelet "Python Pandas: et uunnværlig verktøy for å jobbe med data") til å lese den ekstraherte teksten inn i DataFrame og lagre DataFrame til en CSV-tabellfil eller XLXS:

Jeg trenger kode som konverterer den resulterende tabellen fra en PDF -fil til en DataFrame. Legg også til kode for å lagre DataFrame til en CSV -fil ↴

💡 LLMs svar :



```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

Run in IDE

Fig. 4.1-4 Konvertering av den uthentede tabellen fra PDF til DataFrame og lagring av tabellen i en CSV -fil.

Hvis det oppstår en feil under kjøring av koden (fig. 4.1-3, fig. 4.1-4) - f.eks. på grunn av manglende biblioteker eller feil filbane - kan feilteksten enkelt kopieres sammen med kildekoden og sendes inn til LLM -modellen. Modellen analyserer feilmeldingen, forklarer hva som er problemet og foreslår løsninger eller ytterligere trinn.

Dermed blir interaksjonen med AI LLM en komplett syklus→→test→ tilbakemelding→ korrigering - uten behov for dyp teknisk kunnskap.

Ved hjelp av en ren tekstforespørsel i LLM chat og et dusin linjer med Python som vi kan kjøre lokalt i en hvilken som helst IDE, konverterte vi et PDF -dokument til et CSV-tabellformat, som i motsetning til et PDF-dokument er lett maskinlesbart og raskt kan integreres i et hvilket som helst datahåndteringssystem.

Vi kan bruke denne koden (figur 4.1-3, figur 4.1-4) ved å kopiere den fra et hvilket som helst LLM-chatterom til titusener eller tusenvis av nye PDF-dokumenter på serveren, og på den måten automatisere prosessen med å konvertere en strøm av ustrukturerte dokumenter til et strukturert CSV-tabellformat.

Men PDF-dokumenter inneholder ikke alltid tekst, som oftest er det skannede dokumenter som må behandles som bilder. Selv om bilder i seg selv er ustrukturerte, gjør utviklingen og anvendelsen av gjenkjenningsbiblioteker det mulig for oss å trekke ut, behandle og analysere innholdet, slik at vi kan utnytte disse dataene fullt ut i forretningsprosessene våre.

Konverter JPEG- og PNG-bilder til strukturert form

Bilder er en av de vanligste formene for ustrukturerte data. I bygg- og anleggsbransjen og mange andre bransjer lagres store mengder informasjon i form av skannede dokumenter, skjemaer, fotografier og tegninger. Slike data inneholder verdifull informasjon, men kan ikke bearbeides direkte, for eksempel i et Excel-regneark eller en database. Bilder inneholder mye kompleks informasjon fordi innholdet, fargene og teksturene varierer, og det kreves spesialbehandling for å trekke ut nyttig informasjon.

Problemet med å bruke bilder som datakilde er mangelen på struktur. Bilder formidler ikke mening på en direkte, lett kvantifiserbar måte som en datamaskin umiddelbart kan forstå eller behandle, slik et Excel-regneark eller en databasetabell gjør. For å konvertere ustrukturerte bildedata til en strukturert form må man bruke spesielle biblioteker som er i stand til å tolke den visuelle informasjonen de inneholder (figur 4.1-5).

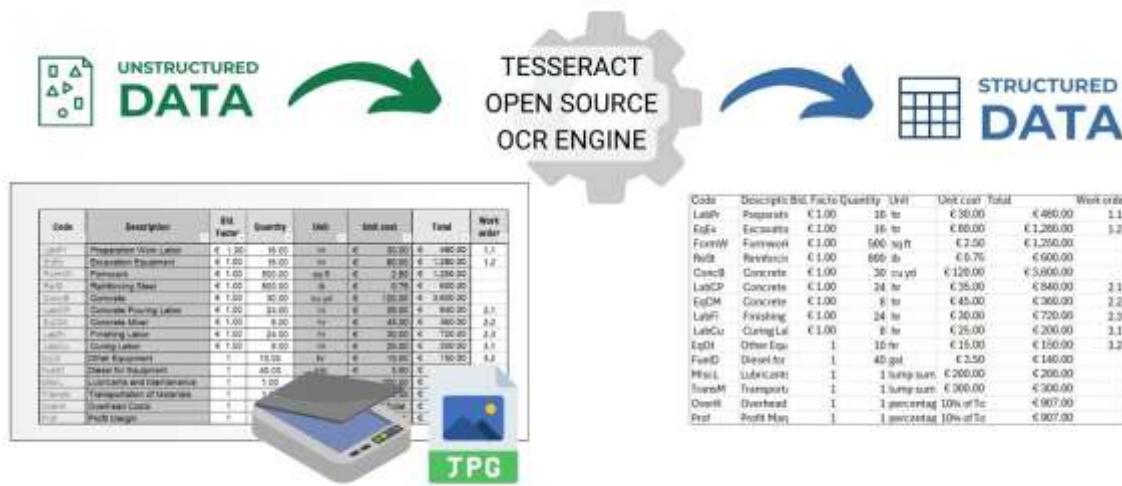


Fig. 4.1-5 Konvertering av skannede dokumenter og bilder til strukturerte formater er mulig med spesielle OCR-verktøy.

OCR-teknologi (Optical Character Recognition) brukes til å trekke ut tekst fra bilder. Den gjør det mulig å gjenkjenne bokstaver og tall i skannede dokumenter, bilder og PDF -filer, og gjøre dem om til redigerbar og maskinlesbar tekst. OCR-teknologier har lenge vært brukt i dokumentautomatisering, og i dag kan de enkelt integreres i alle forretningsprosesser og Python -applikasjoner. Et av de mest populære OCR-verktøyene er

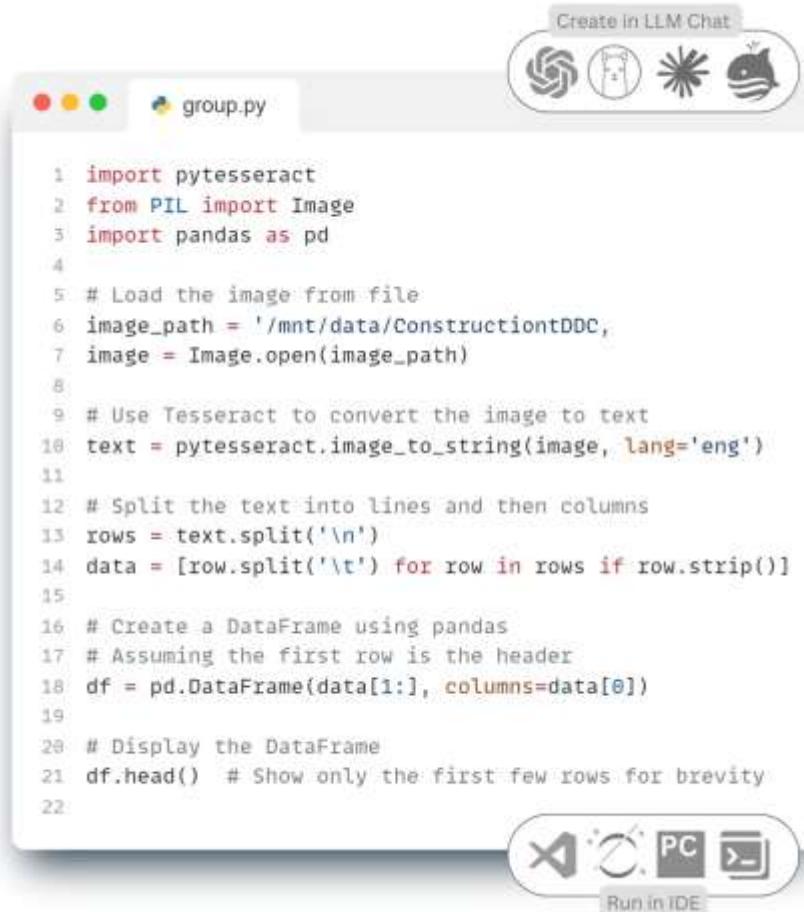
Tesseract, en åpen kildekode, som opprinnelig ble utviklet av HP™ og nå støttes av Google™. Det støtter over 100 språk og har høy gjenkjennelsesnøyaktighet.

La oss be LLM chat om å skrive et kodeeksempel for å hente data fra en skannet eller fotografert tabell på en strukturert måte.

- ⌚ Send en tekstforespørrelse til LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller andre):

Skriv kode for å konvertere et JPEG-bilde som inneholder en tabell til en DataFrame-tabell ↴

- ⌚ LLMs svar vil i de fleste tilfeller foreslå å bruke Pytesseract-biblioteket til å gjenkjenne tekst i bilder :



```

Create in LLM Chat
group.py

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC,
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22
Run in IDE

```

Fig. 4.1-6 Konverter tekst hentet fra en tabell med bilder eller fotografier til en strukturert tabellfremstilling.

I dette eksemplet - koden (figur 4.1-6) hentet fra LLM, bruker pytesseract-biblioteket (Tesseract for Python)

til å konvertere et bilde til tekst ved hjelp av OCR (optisk tegngjenkjenning) og Pandas-biblioteket til å konvertere denne teksten til en strukturert form, dvs. en DataFrame.

Konverteringsprosessen omfatter vanligvis forbehandling for å forbedre bildekvaliteten, og deretter brukes ulike algoritmer for mønstergjenkjenning, funksjonsekstraksjon eller objektgjenkjenning. Resultatet er at ustrukturert visuell informasjon konverteres til strukturerte data.

Selv om PDF og bilder er viktige kilder til ustrukturert informasjon, er den virkelige -mesteren når det gjelder volum tekst generert i e-poster, chatter, møter og messengers. Disse dataene er ikke bare mange - de er også spredte, uformelle og ekstremt dårlig strukturert.

Konvertering av tekstdata til en strukturert form

I tillegg til PDF-dokumenter med tabeller (fig. 4.1-2) og skannede versjoner av tabellskjemaer (fig. 4.1-5), er en betydelig del av informasjonen i prosjektdokumentasjonen presentert i tekstform. Det kan være både sammenhengende setninger i tekstdokumenter og fragmentariske registreringer spredt over tegninger og skjemaer. Under moderne databehandlingsforhold er en av de vanligste oppgavene å konvertere slik tekst til et strukturert format som egner seg for analyse, visualisering og beslutningstaking.

Sentralt i denne prosessen står taksonomi, et klassifiseringssystem som gjør det mulig å organisere informasjon i kategorier og underkategorier basert på fellestrekke.

En taksonomi er en hierarkisk klassifiseringsstruktur som brukes til å gruppere og organisere objekter. I tekstbehandlingssammenheng fungerer den som et grunnlag for systematisk kategorisering av elementer i semantiske kategorier, noe som forenkler analysen og forbedrer kvaliteten på databehandlingen.

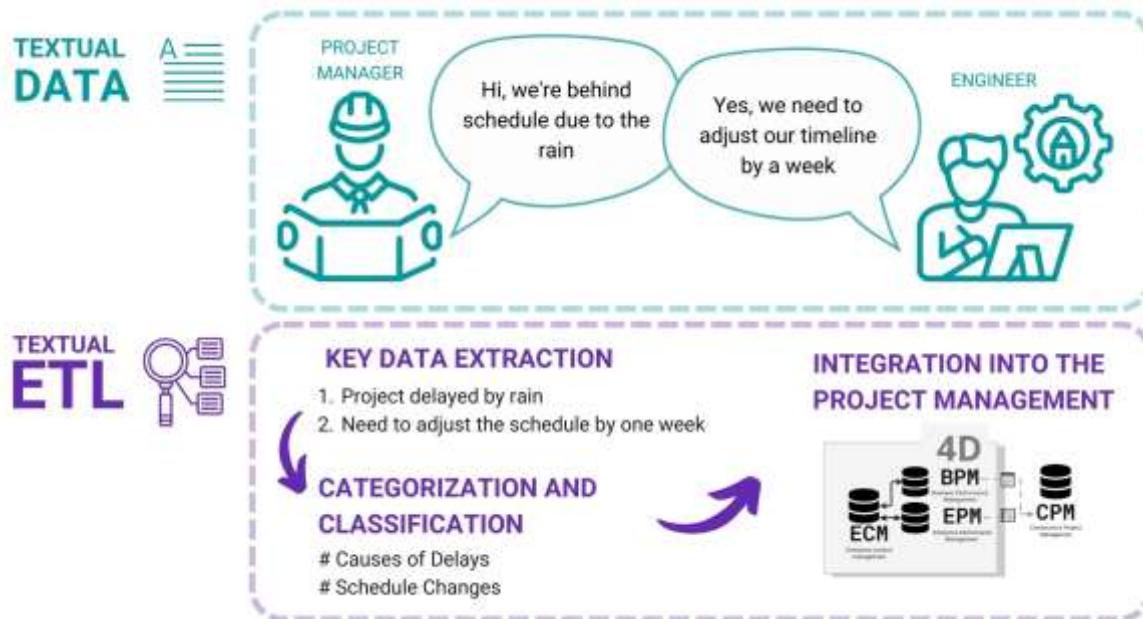
Taksonomisk opprettelse ledsages av trinnene utvinning av entiteter, kategorisering og kontekstualisering. For å modellere prosessen med å trekke ut informasjon fra tekstdata må vi følge følgende trinn, som ligner på dem vi allerede har brukt for å strukturere data fra PDF-dokumenter:

- **Utdrag:** Du må analysere tekstdata for å hente ut informasjon om forsinkelser og endringer i prosjektplanen.
- **Kategorisering og klassifisering:** Kategoriserer mottatt informasjon, f.eks. årsaker til forsinkelser og endringer i tidsplanen.
- **Integrasjon (Load):** Til slutt klargjør vi strukturerte data for integrasjon i eksterne datahåndteringssystemer.

Tenk deg en situasjon: Vi har en dialog mellom en prosjektleder og en ingeniør som diskuterer problemer med tidsplanen. Målet vårt er å trekke ut de viktigste elementene (årsaker til forsinkelser, justeringer av tidsplanen) og presentere dem på en strukturert måte (figur 4.1-7).

La oss utføre uttrekket basert på de forventede nøkkelordene, opprette en DataFrame for å simulere datauttrekk og etter transformasjon en ny DataFrame-tabell som vil inneholde kolonner for dato,

hendelse (f.eks. årsak til forsinkelse) og handling (f.eks. endring av tidsplan).



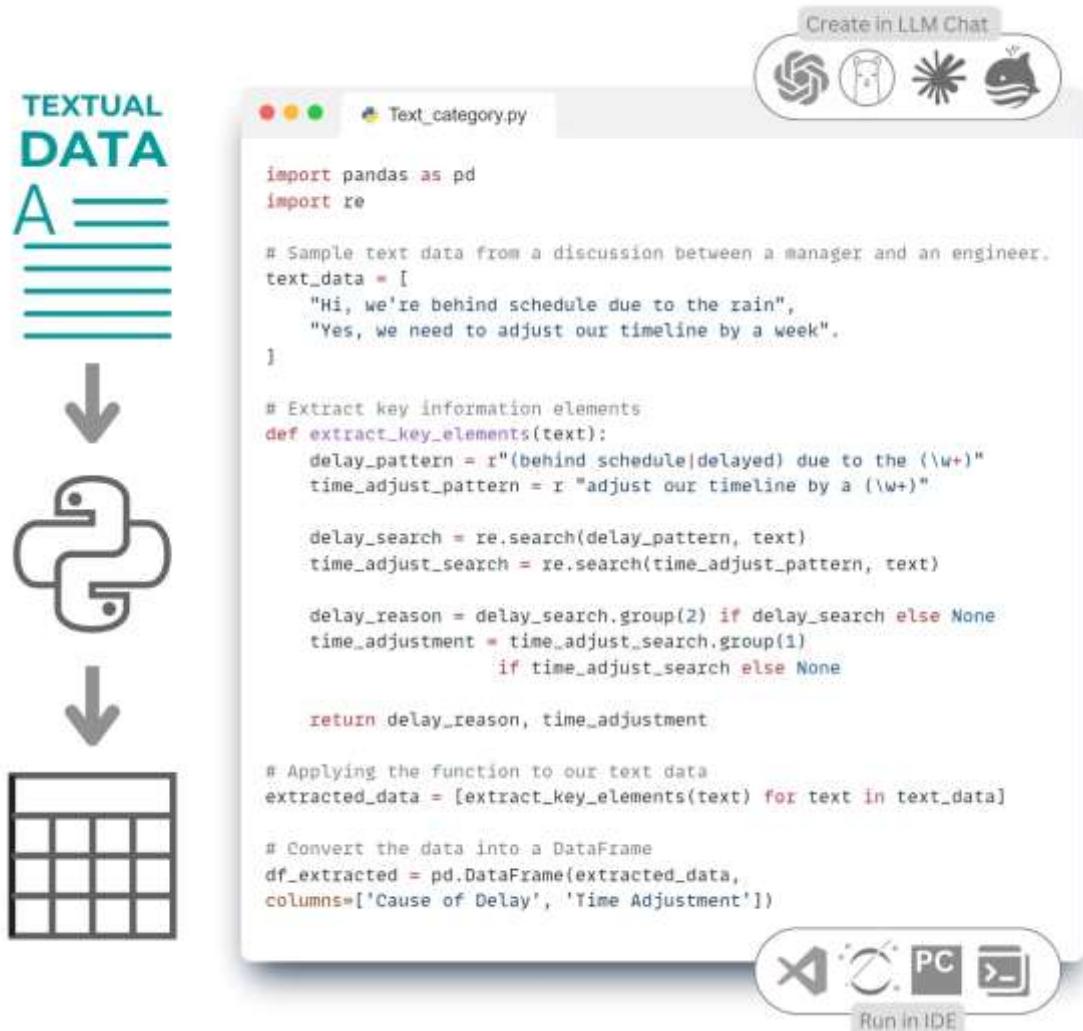
Figur 4.1-7 Utheving av nøkkelinformasjon fra teksten om behovet for å justere tidslinjer og integrere endringer i prosjektstyringssystemet.

Her er koden for å løse problemet ved hjelp av et tekstspørsmål i en av språkmodellene som i de foregående eksemplene.

- Send en tekstforespørsel til et hvilket som helst LLM chatterom:

Jeg har en samtale mellom en leder: "Hallo, vi er forsinket på grunn av regn" og en ingeniør: "Ja, vi må justere tidsfristen med en uke". Jeg trenger et skript som analyserer fremtidige lignende tekstdialoger, trekker ut årsakene til forsinkelser og nødvendige fristjusteringer fra dem, og deretter genererer en DataFrame fra disse dataene. DataFrame skal deretter lagres i en CSV -fil ↵

- Svaret fra LLM vil typisk inneholde Python -kode som bruker regulære uttrykk (re - Regex) og Pandas (pd)-biblioteket :



Figur 4.1-8 Utheving av nøkkelinformasjon fra teksten om behovet for å justere tidsfrister i en tabell.

I dette eksempelet (figur 4.1-7) analyseres tekstdata som inneholder korrespondanse mellom en prosjektleder og en ingeniør, for å identifisere og trekke ut spesifikk informasjon som kan påvirke styringen av fremtidige prosjekter med lignende dialoger. Ved hjelp av regulære uttrykk (vi kommer nærmere inn på regulære uttrykk i kapittelet "Strukturerte krav og regulære RegEx-uttrykk") identifiseres årsakene til prosjektforsinkelser og nødvendige justeringer av tidsplanen ved hjelp av mønstre. Funksjonen som er skrevet i dette eksemplet, trekker ut enten årsaken til forsinkelsen eller tidsjusteringen fra strengene basert på mønstrene: ved å velge ordet etter "because of" som årsak til forsinkelsen eller ordet etter "by" som tidsjustering.

Hvis en rad nevner en forsinkelse på grunn av været, blir "regn" identifisert som årsaken; hvis en rad nevner

en tidsjustering for en bestemt periode, blir denne perioden trukket ut som en tidsjustering (figur 4.1-9). Hvis noen av disse ordene mangler i en rad, får den tilsvarende attributkolonnen verdien "Ingen".

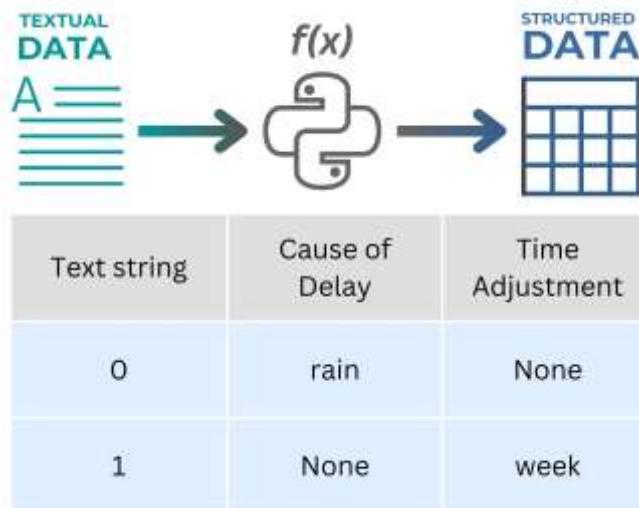


Fig. 4.1-9 Oppsummeringstabellen som fås som DataFrame, etter at koden er kjørt, inneholder informasjon om forsinkelser og nødvendige tidsjusteringer.

Strukturering og parametrisering av forhold fra teksten (dialog, brev, dokument) gjør det mulig å raskt eliminere forsinkelser i konstruksjonen: for eksempel kan mangel på arbeidere påvirke arbeidstempoet i dårlig vær, slik at selskaper som kjenner forsinkelsesparametrene fra dialogene (fig. 4.1-9) mellom formannen på byggeplassen og prosjektlederen - på forhånd kan styrke mannskapet i tilfelle en ugunstig prognose.

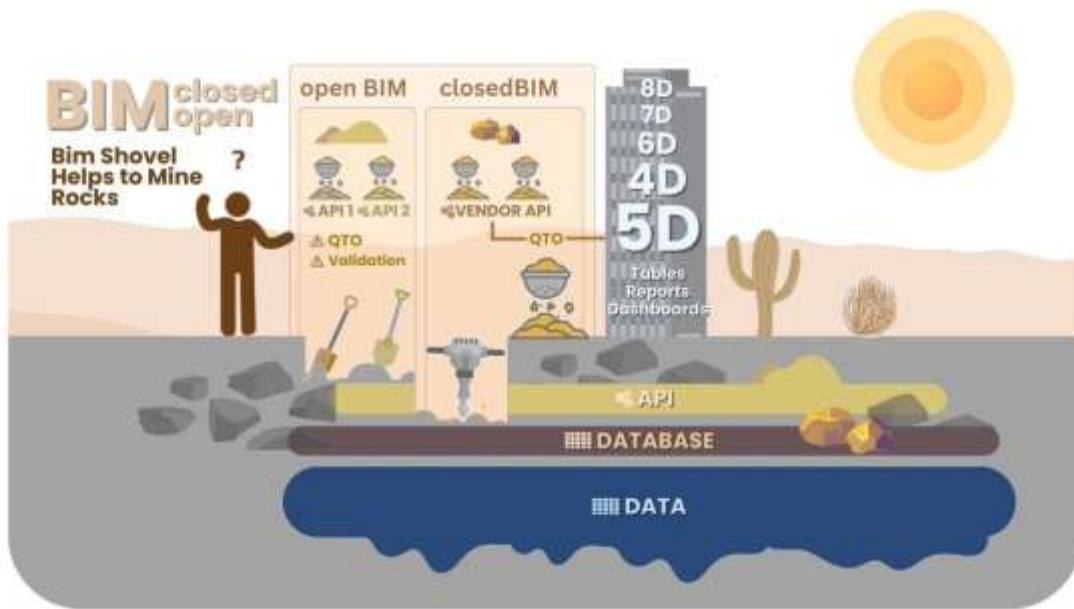
Konverteringen av dokumenter og bilder til et strukturert format kan gjøres ved hjelp av relativt enkle, åpne og gratis kategoriseringsbaserede verktøy.

Kategorisering av elementer er også en sentral del av arbeidet med prosjektdata, spesielt i forbindelse med bruk av CAD-programvare (BIM).

Konvertering av CAD-data (BIM) til en strukturert form

Strukturering og kategorisering av DAK-data (BIM) er en mer kompleks oppgave fordi dataene som lagres i DAK-databaser (BIM) nesten alltid er i lukkede eller komplekse parametriske formater, som ofte kombinerer geometriske dataelementer (semistrukturerte) og metainformasjonselementer (semistrukturerte eller strukturerte data) samtidig.

Native dataformater i CAD (BIM) -systemer er vanligvis beskyttet og utilgjengelige for direkte bruk, med mindre spesialisert programvare eller API - grensesnitt fra utvikleren selv (fig. 4.1-10). Slik dataisolering danner lukkede lagringssiloer som begrenser den frie utvekslingen av informasjon og hemmer etableringen av ende-til-ende digitale prosesser i bedriften.



Figur 4.1-10 CAD-profesjonelle (BIM) kan få tilgang til opprinnelige data via API -tilkoblinger eller leverandørverktøy.

I spesielle CAD (BIM) -formater samles informasjon om prosjektelementenes egenskaper og attributter i et hierarkisk klassifiseringssystem, der enheter med tilsvarende egenskaper er plassert, som frukten på et frukttræ, i de nyeste nodene i -dataklassifiseringsgrenene (fig. 4.1-11).

Det er to måter å hente ut data fra slike hierarkier på: enten manuelt, ved å klikke på hver node, som om man bearbeidet et tre, og hugge ned utvalgte grener av kategorier og typer med en øks. Alternativt kan man bruke programmeringsgrensesnitt (API-er) for å hente ut og gruppere data på en mer effektiv og automatisert måte, slik at de til slutt kan omdannes til en strukturert tabell som kan brukes i andre systemer.

Ulike verktøy som Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC eller åpen kildekode-løsninger, som IfcOpSh eller IFCjs for IFC-formatet, kan brukes til å hente ut strukturerte datababler fra CAD (BIM) - prosjekter.

Moderne dataeksport- og konverteringsverktøy gjør det mulig å forenkle databehandling og -klargjøring ved å dele innholdet i CAD-modeller i to hovedkomponenter: geometriinformasjon og attributdata (fig. 4.1-13) - metainformasjon som beskriver egenskapene til designelementene (fig. 3.1-16). Disse to datalagene er knyttet sammen gjennom unike identifikatorer, noe som gjør det mulig å tilordne hvert element med geometribeskrivelse (via parametere eller polygoner) til dets attributter: navn, materiale, fullføringsstadium, kostnad og så videre. Denne tilnærmingen sikrer modellens integritet og gjør det mulig å bruke dataene fleksibelt både til visualisering (geometriske modelldata) og til analyse- eller administrasjonsoppgaver (strukturert eller løst strukturert), ved å arbeide med de to datatypene hver for seg eller parallelt.

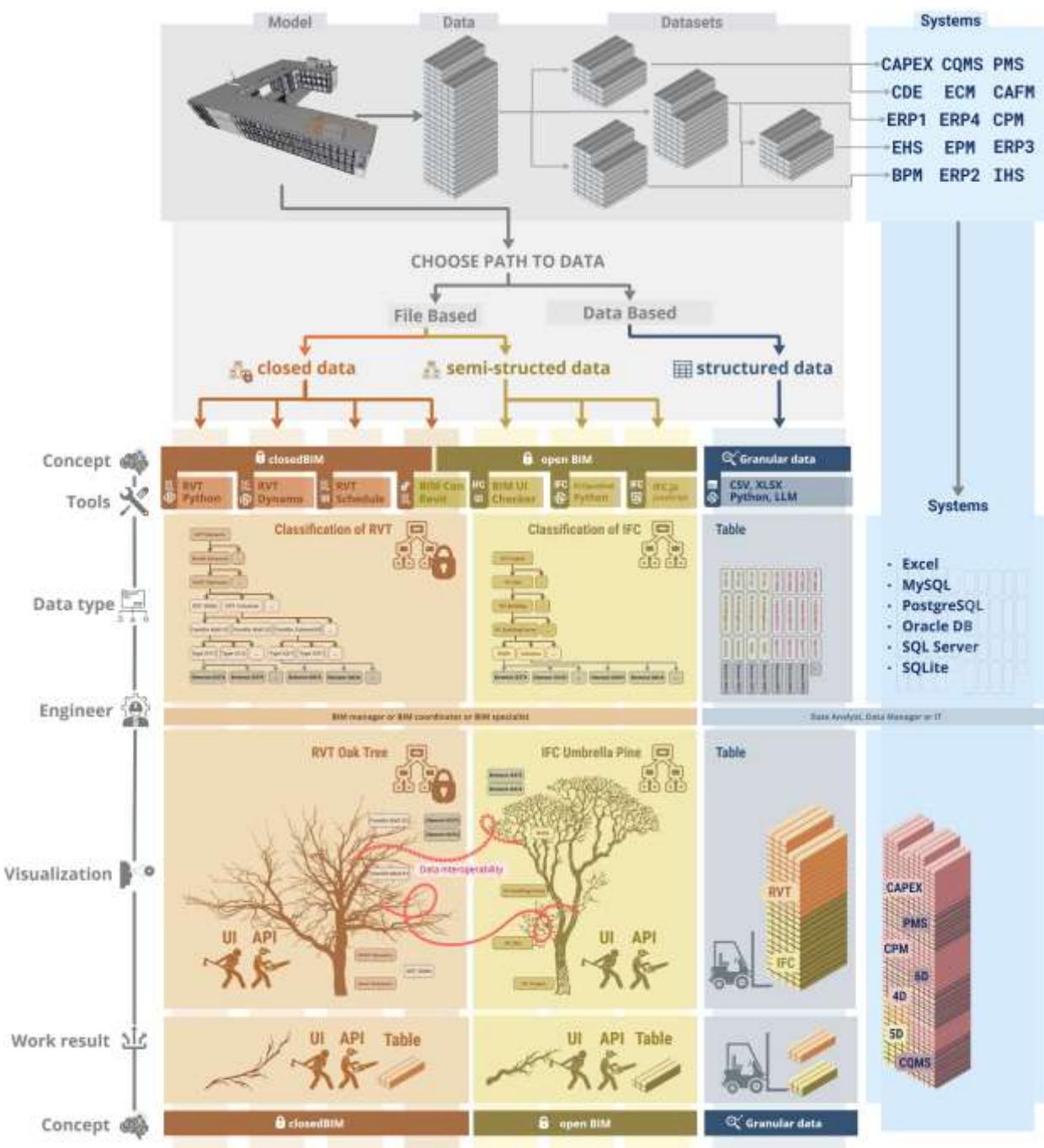


Fig. 4.1-11 Visningen av informasjon fra CAD-databaser (BIM) presenteres for brukeren i form av klassifiseringstrær.

Med utviklingen av reverse engineering-teknologier og fremveksten av SDK-er (Software Development Kit) for konvertering av DAK-data - har tilgjengeligheten og konverteringen av data fra lukkede DAK-programvareformater (BIM) blitt mye enklere. Det er nå mulig å konvertere data fra lukkede formater til universelle formater som egner seg for analyse og bruk i andre systemer på en lovlig og sikker måte. Historien om de første verktøyene for omvendt konstruksjon ("Open DWG") og kampen om dominans over

CAD-leverandørenes formater ble diskutert i kapittelet "Strukturerte data: grunnlaget for digital transformasjon".

Verktøy for omvendt konstruksjon gjør det mulig å hente data fra lukkede, proprietære formater og bryte ned informasjon fra det blandede CAD (BIM) -formatet til de datatypene og -formatene som brukeren trenger, slik at det blir enklere å behandle og analysere dem.

Ved hjelp av reverse engineering og direkte tilgang til informasjon fra CAD-databaser gjør informasjon tilgjengelig, noe som muliggjør åpne data og åpne verktøy, samt dataanalyse ved hjelp av standardverktøy, bygging av rapporter, visualiseringer og integrering med andre digitale systemer (fig. 4.1-12).

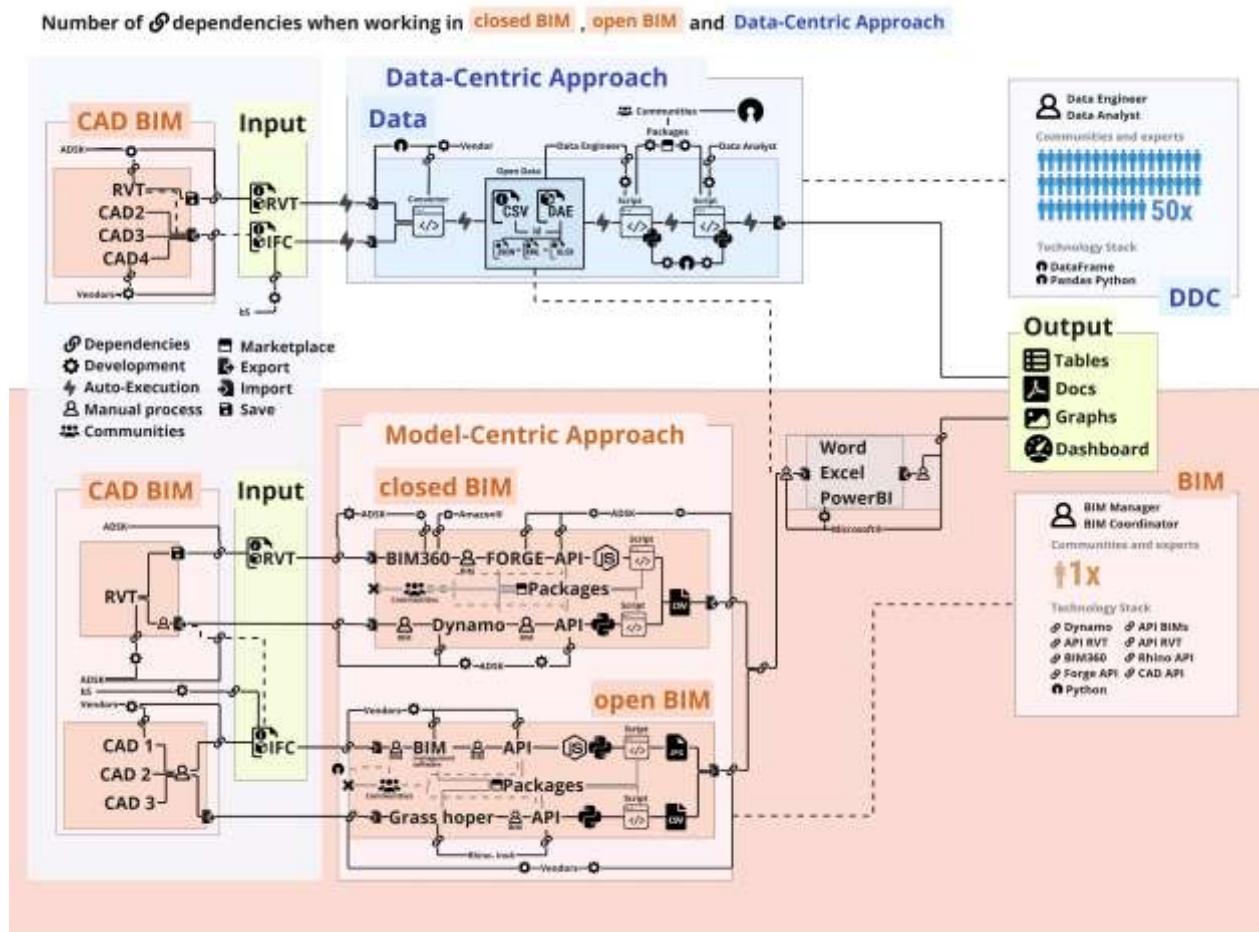


Fig. 4.1-12 Direkte tilgang til DAK-data minimerer avhengigheten av programvareplattformer og går i retning av en datasentrisk tilnærming.

Siden 1996 for DWG-formatet, siden 2008 for DGN-formatet og siden 2018 for RVT har det vært mulig å konvertere opprinnelig lukkede CAD-dataformater til alle andre formater, inkludert strukturerte formater, på en enkel og effektiv måte ved hjelp av verktøy for omvendt konstruksjon (fig. 4.1-13). I dag bruker nesten alle store CAD (BIM) og store ingeniørfirmaer i verden SDK-er - reverse engineering-verktøy for å hente ut data fra lukkede CAD (BIM)-leverandørformater [92].



Fig. 4.1-13 Ved hjelp av verktøy for omvendt konstruksjon kan CAD (BIM) -programdatabaser konverteres til en hvilken som helst praktisk datamodell.

Konvertering av data fra lukkede, proprietære formater til åpne formater og separering av blandede CAD (BIM) formater i geometriske og metainformasjonsattributtdata forenkler arbeidet med dataene, og gjør dem tilgjengelige for analyse, manipulering og integrering med andre systemer (fig. 4.1-14).

I dagens arbeid med DAK-data (BIM) har vi nådd et punkt der vi ikke trenger å be om tillatelse fra DAK-leverandører (BIM-leverandører) for å få tilgang til informasjon fra DAK-formater.

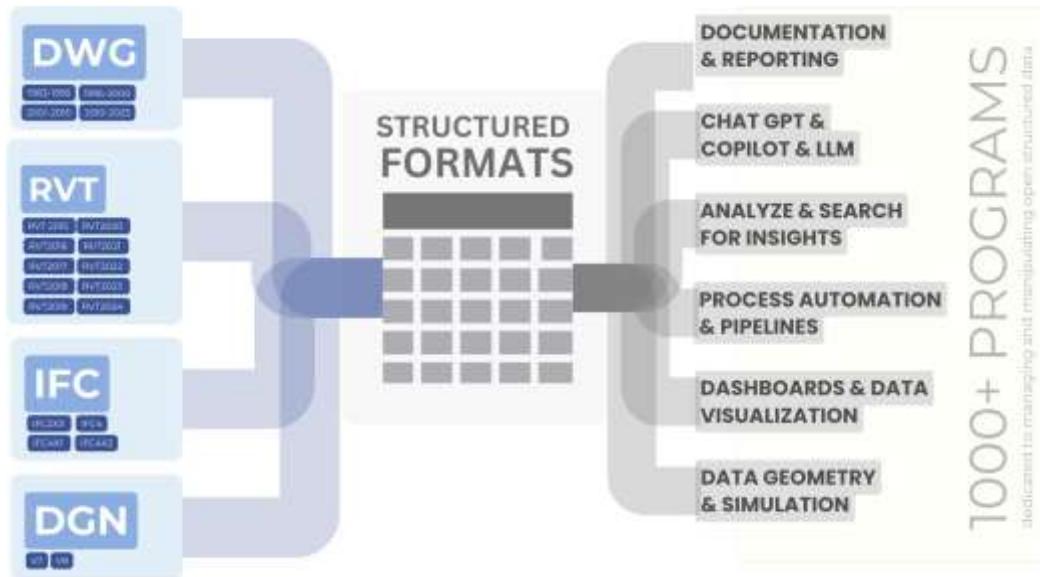


Fig. 4.1-14 Moderne SDK-verktøy tillater lovlig konvertering av data fra proprietære CAD-databaseformater (BIM).

Dagens trender innen behandling av CAD-designdata fortsetter å bli formet av viktige markedsaktører - CAD-leverandører som arbeider for å styrke sin posisjon i dataverdenen og skape nye formater og konsepter.

CAD - leverandørene av løsninger går i retning av strukturerte data

Fra og med 2024 gjennomgår design- og byggebransjen et betydelig teknologisk skifte når det gjelder bruk

og behandling av data. I stedet for fri tilgang til designdata fokuserer leverandørene av CAD -systemer på å fremme de neste nye konseptene. Tilnærmingen som BIM (opprettet i 2002) og åpen BIM (opprettet i 2012) viker gradvis plassen for moderne teknologiske løsninger som DAK-leverandørene begynner å promotere [93]:

- Overgang til bruk av "granulære" data som muliggjør effektiv informasjonshåndtering og et skifte i retning av dataanalyse
- USD -format og implementering av Entity-component-system-tilnærming (ECS) for fleksibel dataorganisering
- Aktiv bruk av kunstig intelligens i databehandling, prosessautomatisering og dataanalyse
- Utvikling av interoperabilitet - bedre samspill mellom ulike programmer, systemer og databaser

Hvert av disse aspektene vil bli diskutert mer inngående i den sjette delen av boken "CAD and BIM: marketing, reality and the future of design data in construction". I dette kapittelet vil vi bare kort skissere den generelle endringsvektoren: De store DAK-leverandørene forsøker nå å revurdere måten prosjekteringsinformasjonen er strukturert på. En av de viktigste endringene er at den klassiske filbaserte lagringsmodellen forlates til fordel for en detaljert, analyseorientert dataarkitektur som gir kontinuerlig tilgang til individuelle modellkomponenter [93].

Det som skjer, er at bransjen faser ut tungvinte, spesialiserte og parametriske formater som krever geometriske kjerner, til fordel for mer universelle, maskinlesbare og fleksible løsninger.

En slik drivkraft for endring er USD-formatet (Universal Scene Description), som opprinnelig ble utviklet i datagrafikkbransjen, men som allerede er anerkjent i tekniske anvendelser takket være utviklingen av NVIDIA Omniverse (og Isaac Sim)-plattformen for simuleringer og visualiseringer [93]. I motsetning til den parametriske IFC har USD en enklere struktur og gjør det mulig å beskrive geometri og objektekvenskaper i JSON-format (figur 4.1-15), noe som forenkler informasjonsbehandling og gjør det raskere å integrere den i digitale prosesser. Det nye formatet gjør det mulig å lagre geometri (i tillegg til BREP -NURBS - flere detaljer i del 6 av boken) i form av MESH-polygoner, og objektekvenskaper i JSON, noe som gjør det mer praktisk for automatiserte prosesser og arbeid i skyøkosystemer [94].

Noen CAD- og ERP-leverandører bruker allerede lignende formater (f.eks. NWD, SVF, CP2, CPIXML), men de fleste av dem er fortsatt lukkede og utilgjengelige for ekstern bruk, noe som begrenser mulighetene for dataintegrasjon og gjenbruk. I denne sammenhengen kan USD spille samme rolle som DXF gjorde i sin tid - et åpent alternativ til proprietære formater som DWG.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	b5 (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in b5 (IAI) since the introduction

Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

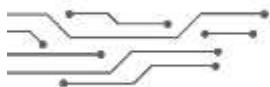
Fig. 4.1-15 USD formatet som et forsøk fra CAD-leverandørene på å imøtekommе kravet om interoperabilitet og uavhengighet av designdata fra geometriske kjerner.

De store utviklernes overgang til åpne og forenklede USD, GLTF, OBJ, XML (lukket NWD, CP2, SVF, SVF2, CPIXML) og lignende formater (fig. 3.1-17) gjenspeiler den globale trenden og bransjens krav om forenkling av data og økt tilgjengelighet. I de kommende årene kan vi forvente en gradvis overgang fra komplekse parametriske standarder og formater som er avhengige av geometriske kjerner, til lettere og mer strukturerte løsninger. Denne overgangen vil akselerere digitaliseringen av byggebransjen, legge til rette for automatisering av prosesser og forenkle datautvekslingen.

Til tross for de strategiske planene til CAD -leverandørene om å fremme nye åpne formater, kan fagfolk i byggebransjen også få full tilgang til data fra lukkede CAD-systemer, uten behov for CAD-verktøy (BIM), ved å bruke reverse engineering-verktøy.

Alle disse trendene fører uunngåelig til et skifte fra store, monolittiske 3D-modeller til universelle, strukturerte data og til bruk av formater som lenge har vært utprøvd i andre bransjer. Når prosjektteamene begynner å se på DAK-modeller ikke bare som visuelle objekter eller et sett med filer, men som databaser som inneholder kunnskap og informasjon, endres tilnærmingen til design og administrasjon dramatisk.

Når teamene har lært seg å trekke ut strukturerte data fra dokumenter, tekster, tegninger og CAD-modeller, og har tilgang til databaser, er det neste viktige trinnet datamodellering og kvalitetssikring. Det er dette trinnet som i stor grad avgjør hvor raskt det går å behandle og transformere informasjonen som til slutt skal brukes til å ta beslutninger i spesifikke oppgaver.



KAPITTEL 4.2.

KLASSIFISERING OG INTEGRERING: ET FELLES SPRÅK FOR BYGGDATA

Hvor raskt beslutningene tas, avhenger av kvaliteten på dataene

Dagens arkitektur for designdata er i ferd med å gjennomgå grunnleggende endringer. Bransjen beveger seg bort fra voluminøse, isolerte modeller og lukkede formater og over mot mer fleksible, maskinlesbare strukturer med fokus på analyse, integrering og prosessautomatisering. Overgangen til nye formater alene er imidlertid ingen garanti for effektivitet - kvaliteten på selve dataene står uunngåelig i sentrum.

I denne boken snakker vi mye om formater, systemer og prosesser. Men alle disse tiltakene er meningsløse uten ett nøkkelement: data som er til å stole på. Datakvalitet er hjørnestenen i digitaliseringen, og vi kommer tilbake til det i de neste delene av boken.

Moderne byggefirmaer - spesielt de store - bruker dusinvis og noen ganger tusenvis av forskjellige systemer og databaser (fig. 4.2-1). Disse systemene må ikke bare fylles med ny informasjon med jevne mellomrom, men de må også samhandle effektivt med hverandre. Alle nye data som genereres som et resultat av behandlingen av innkommende informasjon, integreres i disse miljøene og brukes til å løse spesifikke forretningsoppgaver.

Og mens beslutninger om spesifikke forretningsoppgaver tidligere ble tatt av toppledere - såkalte HiPPOer (fig. 2.1-9) - på grunnlag av erfaring og intuisjon, er denne tilnærmingen i dag, med den kraftige økningen i informasjonsmengden, i ferd med å bli kontroversiell. Automatiserte analyser, som arbeider med sanntidsdata, er i ferd med å erstatte den.

"Tradisjonelle, manuelle" diskusjoner om forretningsprosesser på ledernivå vil gå over til operative analyser, som krever raske svar på forretningsspørsmål.

Tiden da regnskapsførere, formenn og kalkulatører manuelt genererte rapporter, oppsummeringstabeller og utstillingsvinduer med prosjektdata over dager og uker, hører fortiden til. I dag er raske og tidsriktige beslutninger i ferd med å bli en nøkkelfaktor for å oppnå konkurransefortrinn.

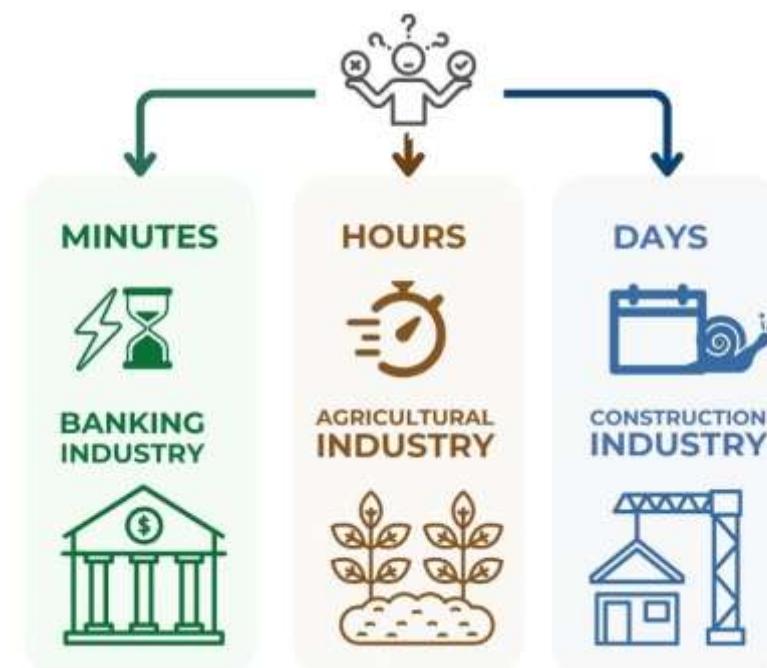


Fig. 4.2-1 I byggebransjen tar det flere dager å beregne og ta beslutninger, i motsetning til andre bransjer der dette skjer i løpet av timer eller minutter.

Den største forskjellen mellom byggebransjen og mer digitalt avanserte bransjer (figur 4.2-1) er den lave datakvaliteten og standardiseringen. Utdaterte metoder for generering, overføring og behandling av informasjon forsinkes prosesser og skaper kaos. Mangelen på enhetlige standarder for datakvalitet hindrer implementeringen av.

En av hovedutfordringene er fortsatt den dårlige kvaliteten på inngangadataene, samt mangelen på formaliserte prosesser for utarbeidelse og validering av dem. Uten pålitelige og konsistente data er det umulig å få til en effektiv integrering mellom systemene. Dette fører til forsinkelser, feil og økte kostnader i alle faser av prosjektets livssyklus.

I de følgende avsnittene i boken ser vi nærmere på hvordan du kan forbedre datakvaliteten, standardisere prosesser og forkorte veien fra informasjon til validerte og konsistente data av høy kvalitet.

Datastandardisering og -integrasjon

Effektiv datahåndtering krever en tydelig standardiseringsstrategi. Bare med klare krav til datastruktur og -kvalitet kan datavalidering automatiseres, manuelle operasjoner reduseres og informerte beslutninger tas raskere i alle faser av et prosjekt.

I det daglige må et byggefirma behandle hundrevis av filer hver dag: e-post, PDF-dokumenter, CAD-designfiler, -data fra IoT-sensorer, som må integreres i selskapets forretningsprosesser.

Skogen i en bedrifts økosystem av databaser og verktøy (figur 4.2-2) må lære seg å utlede næringsstoffer fra innkommende data i flere formater for å oppnå de resultatene bedriften ønsker.

For å håndtere dataflyten på en effektiv måte trenger du ikke nødvendigvis å ansette en hærskare av ledere, du må først utvikle strenge krav og standarder for dataene og bruke egnede verktøy for automatisk validering, harmonisering og behandling av dem.

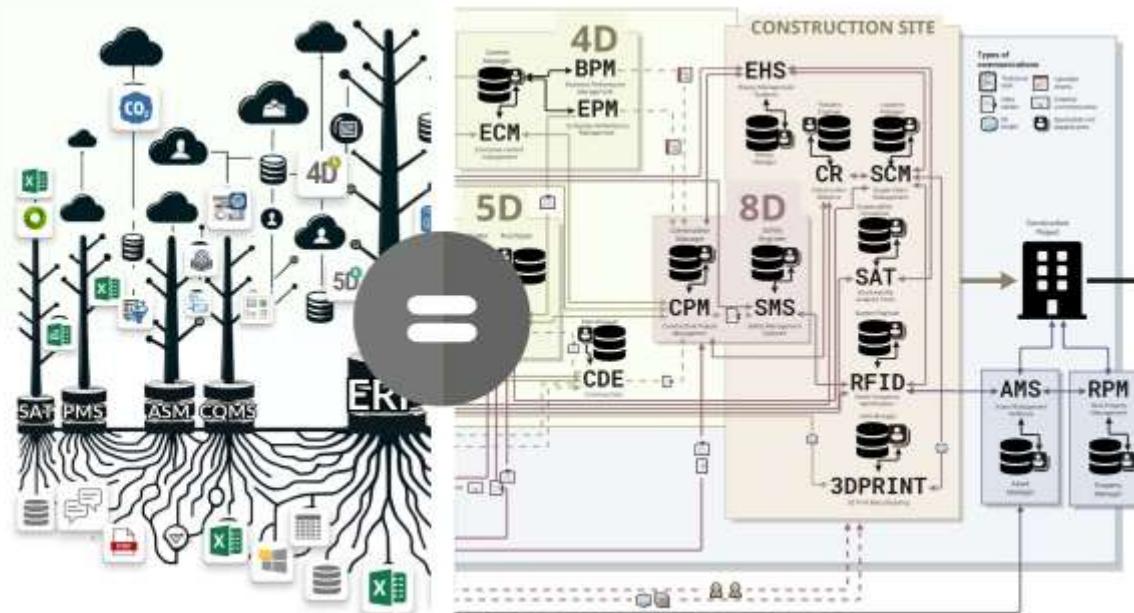


Fig. 4.2-2 For å sikre et sunt økosystem i en bedrift må systemene være av høy kvalitet og ha tilgang til ressurser i tide.

For å kunne automatisere datavaliderings- og harmoniseringsprosessen (for påfølgende integrering) bør du starte med å beskrive de nødvendige minimumskravene til data for hvert enkelt system. Disse kravene definerer:

- Hva trenger du egentlig å få tak i?
- I hvilken form (struktur, format)?
- Hvilke attributter er obligatoriske?
- Hvilke toleranser i nøyaktighet og fullstendighet er akseptable?

Datakravene beskriver kriterier for kvalitet, struktur og fullstendighet i den informasjonen som mottas og behandles. For eksempel er det viktig at tekster i PDF-dokumenter er formatert nøyaktig i henhold til bransjestandarder (fig. 7.2-14 - fig. 7.2-16). Objekter i CAD-modeller må ha korrekte attributter (dimensjoner, koder, koblinger til klassifikatorer) (fig. 7.3-9, fig. 7.3-10). Og for kontraktsskanninger er det viktig med tydelige datoer og muligheten til å automatisk trekke ut beløp og nøkkelvilkår (fig. 4.1-7 - fig. 4.1-10).

Formulering av datakrav og automatisk kontroll av samsvar er et av de mest tidkrevende, men kritiske trinnene. Det er det mest tidkrevende trinnet i forretningsprosessene.

Som nevnt i del 3 av denne boken, går mellom 50% og 90% av tiden til business intelligence (BI)-eksperter med til å forberede data i stedet for å analysere dem (figur 3.2-5). Denne prosessen omfatter datainnsamling, verifisering, validering, harmonisering og strukturering.

Ifølge en undersøkelse fra 2016 [95] oppga dataforskere fra en rekke ulike fagfelt at de bruker mesteparten av arbeidstiden (ca. 80%) på det de liker minst å gjøre (figur 4.2-3): å samle inn eksisterende datasett og organisere (forene, strukturere) dem. Dermed blir mindre enn 20 prosent av tiden deres igjen til kreative oppgaver, som å finne mønstre og regelmessigheter som kan føre til ny innsikt og nye oppdagelser.

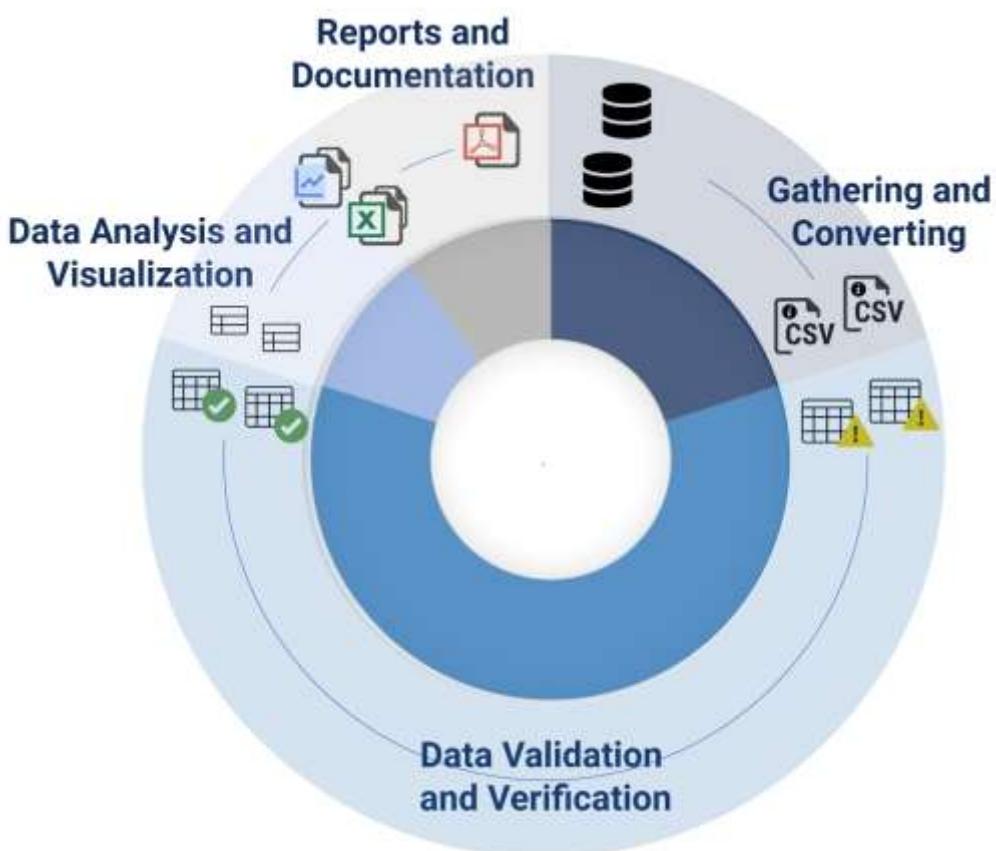


Fig. 4.2-3 Verifisering og sikring av datakvalitet er det mest kostbare, tidkrevende og komplekse trinnet i forberedelsen av data for integrering i andre systemer.

Vellykket datahåndtering i et byggefirma krever en helhetlig tilnærming som omfatter parametrisering av oppgaver, formulering av krav til datakvalitet og bruk av egnede verktøy for automatisert validering.

Digital interoperabilitet starter med krav

I takt med at antallet digitale systemer i bedrifter vokser, øker også behovet for datakonsistens mellom dem. Ledere med ansvar for ulike IT-systemer opplever ofte at de ikke klarer å holde tritt med den økende informasjonsmengden og de mange ulike formatene. I slike tilfeller blir de nødt til å be spesialister om å skape data i en form som egner seg for bruk i andre applikasjoner og plattformer.

Dette krever i sin tur at ingeniører og datagenereringsmedarbeidere må tilpasse seg en rekke ulike krav, ofte uten åpenhet og en klar forståelse av hvor og hvordan dataene skal brukes i fremtiden. Mangelen på standardiserte tilnæringer til håndtering av informasjon fører til ineffektivitet og økte kostnader i verifiseringsfasen, som ofte er manuell på grunn av dataenes kompleksitet og ikke-standardiserte natur.

Problemet med datastandardisering er ikke bare et spørsmål om bekvemmelighet eller automatisering. Det er et direkte økonomisk tap. Ifølge en IBM-rapport fra 2016 er det årlige tapet som følge av dårlig datakvalitet i USA på 3,1 billioner dollar [96]. I tillegg viser studier fra MIT og andre analytiske konsulenterselskaper at kostnadene ved dårlig datakvalitet kan utgjøre så mye som 15-25% av et selskaps inntekter [97].

Under disse forholdene blir det avgjørende å ha klart definerte datakrav og beskrivelser av hvilke parametere, i hvilket format og med hvilken detaljeringsgrad som skal inngå i de opprettede objektene. Uten formalisering av disse kravene er det umulig å garantere datakvalitet og kompatibilitet mellom systemer og prosjektfaser (fig. 4.2-4).

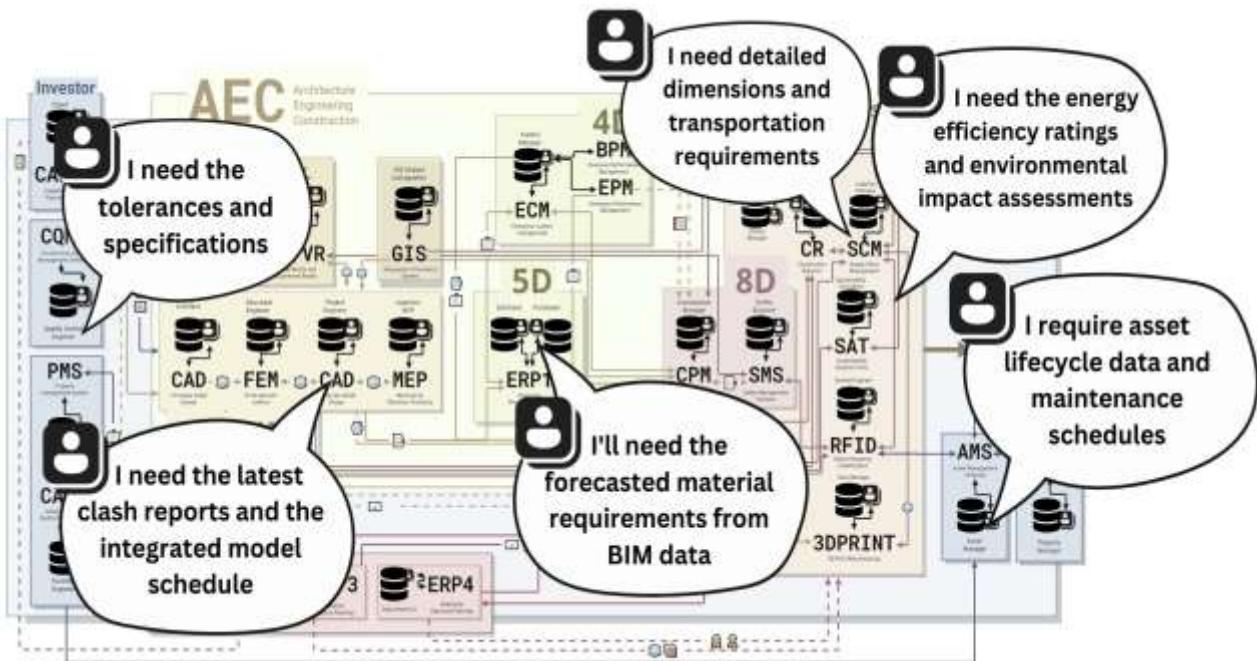
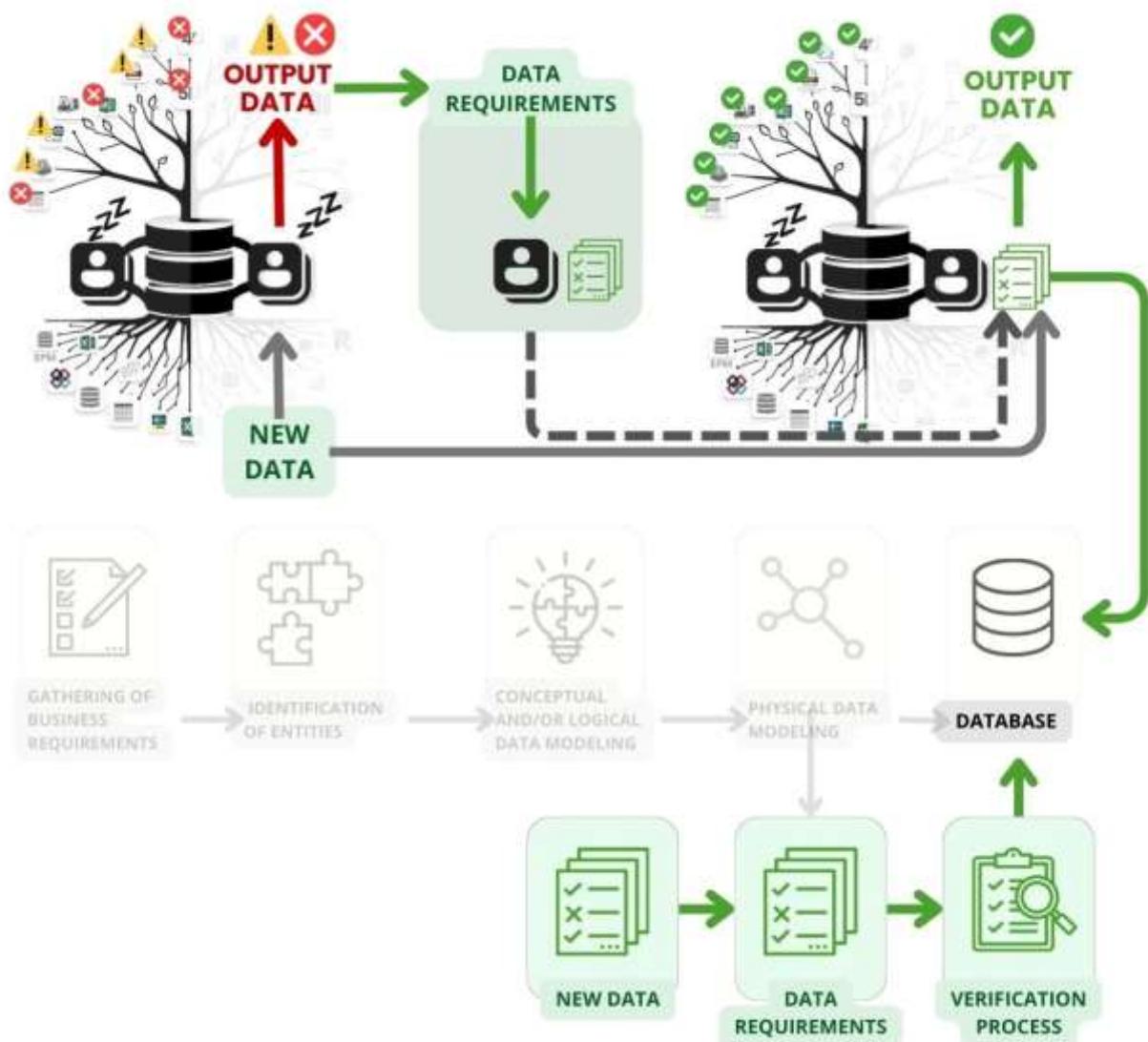


Fig. 4.2-4 Virksomheten er basert på samspillet mellom ulike roller, som hver for seg krever visse parametere og verdier som er avgjørende for å nå virksomhetens mål.

For å kunne formulere de riktige datakravene, må du forstå forretningsprosessene på datanivå. Byggeprosjekter varierer i type, omfang og antall deltakere, og hvert system - enten det dreier seg om modellering (CAD (BIM)), planlegging (ERP 4D), kalkulasjon (ERP 5D) eller logistikk (SCM) - krever sine egne unike parametere for inndata (input entity-elements).

Avhengig av disse behovene må bedriftslederne enten utforme nye datastrukturer for å oppfylle kravene eller tilpasse eksisterende tabeller og databaser. Kvaliteten på dataene som skapes, vil være direkte avhengig av hvor presist og korrekt kravene er formulert (fig. 4.2-5).



Figur 4.2-5 Datakvaliteten avhenger av kvaliteten på kravene som stilles til spesifikke databrukstilfeller.

Siden hvert system har sine egne spesifikke datakrav, er det første trinnet i formuleringen av generelle krav å kategorisere alle elementer som er involvert i forretningsprosessene. Dette innebærer at objektene må deles inn i klasser og grupper av klasser som svarer til spesifikke systemer eller applikasjonsoppgaver. For hver slik gruppe utvikles det separate krav til datastruktur, attributter og kvalitet.

I praksis står imidlertid implementeringen av denne tilnærmingen overfor en stor utfordring: mangelen på et felles språk for gruppering av data. Ulik klassifisering, dupliserte identifikatorer og inkompatible formater fører til at hvert selskap, hver programvare og til og med hvert prosjekt lager sine egne, isolerte datamodeller og -klasser. Resultatet er et digitalt "Babels tårn", der overføring av informasjon mellom systemer krever flere konverteringer til de riktige datamodellene og -klassene, ofte manuelt. Denne barrieren kan bare overvinnes ved å gå i retning av universelle klassifikatorer og standardiserte kravsett.

Et felles konstruksjonsspråk: Klassifiseringsverktøyenes rolle i den digitale transformasjonen

I forbindelse med digitalisering og automatisering av inspeksjons- og prosesseringsprosesser spiller klassifiseringssystemer en spesiell rolle - en slags "digitale ordbøker" som sikrer ensartethet i beskrivelsen og parametrisingen av objekter. Klassifikatorer danner et "felles språk" som gjør det mulig å gruppere data etter betydning og integrere dem mellom ulike systemer, styringsnivåer og faser i prosjektets livssyklus.

Den mest håndgripelige effekten av klassifikatorer er økonomien i bygningens livssyklus, der det viktigste aspektet er optimalisering av de langsigtede driftskostnadene. Studier viser at driftskostnadene utgjør opptil 80% av de totale eierkostnadene for en bygning, noe som er tre ganger høyere enn de opprinnelige byggekostnadene (fig. 4.2-6) [98]. Dette betyr at beslutningen om fremtidige kostnader i stor grad tas på prosjekteringsstadiet

Derfor bør krav fra driftsingeniører (CAFM, AMS, PMS, RPM) bli utgangspunktet for å generere datakrav i designfasen (fig. 1.2-4). Disse systemene bør ikke ses på som prosjektets siste fase, men som en integrert del av hele prosjektets digitale økosystem, fra konsept til avvikling

En moderne klassifikator er ikke bare et system med koder for gruppering. Det er en mekanisme for gjensidig forståelse mellom arkitekter, ingeniører, kalkulatorer, logistikere, vedlikeholds- og IT-systemer. På samme måte som bilens autopilot må gjenkjenne objekter på veien på en entydig og nøyaktig måte, må digitale bygesystemer og deres brukere tolke samme prosjektelement entydig for ulike systemer via elementklassen.

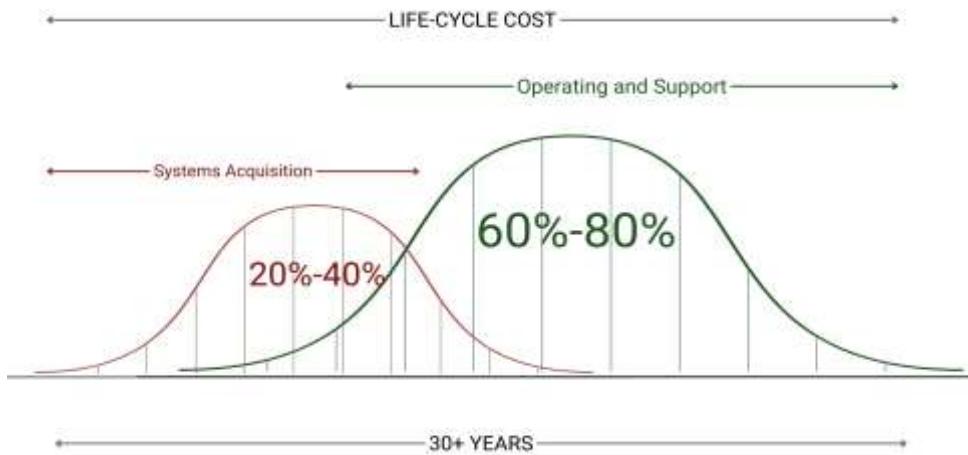


Fig. 4.2-6 Drifts- og vedlikeholdskostnadene er tre ganger så høye som byggekostnadene, og utgjør 60-80% av de totale livssykluskostnadene for en bygning (ifølge [99]).

Graden av utvikling av klassifiseringsverktøy korrelerer direkte med graden av digitalisering og digital modenhet i en bedrift. Organisasjoner med lav digital modenhet står overfor fragmenterte data, inkompatible informasjonssystemer og, som et resultat, inkompatible og ineffektive klassifikatorer. I slike selskaper kan det samme elementet ofte ha ulike grupperingsidentifikatorer i ulike systemer, noe som er et

kritisk hinder for endelig integrering og umuliggjør prosessautomatisering.

For eksempel kan det samme vinduet i et prosjekt merkes forskjellig i CAD modell, kalkulasjons- og vedlikeholdssystem (fig. 4.2-7) på grunn av den flerdimensjonale oppfatningen av elementer hos ulike deltakere i prosessen. For kalkulatøren i vinduskategorien element er volum og kostnad viktig, for vedlikeholdstjenesten - tilgjengelighet og vedlikeholdbarhet, for arkitekten - estetiske og funksjonelle egenskaper. Som et resultat kan det samme elementet kreve forskjellige parametere.

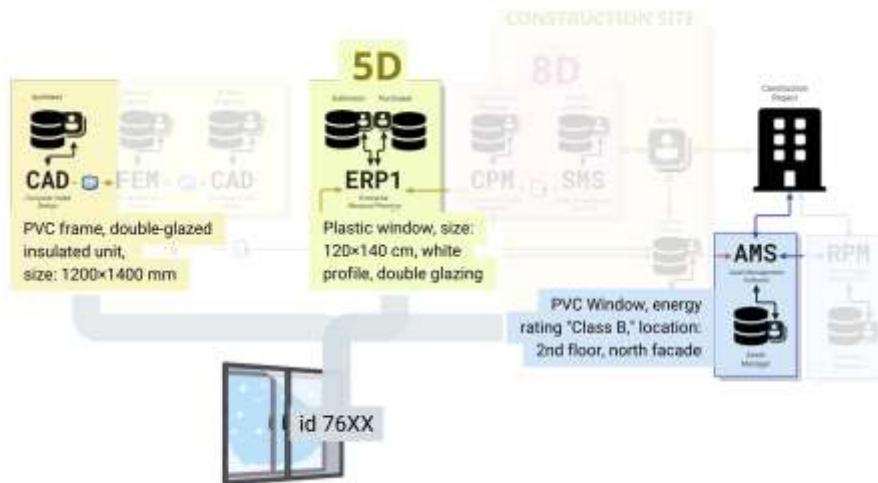


Fig. 4.2-7 Med inkonsekvent klassifisering mellom systemer vil et element miste noe attributtinformasjon ved hvert trinn i overgangen til et annet system.

Fordi det er vanskelig å definere en entydig klassifisering av bygningselementer, tilordner spesialister fra ulike fagområder ofte uforenlig klasse til samme element. Dette fører til at man mister et enhetlig syn på objektet, noe som krever manuell inngrisen for å harmonisere de ulike klassifiseringssystemene og for å etablere konsistens mellom typene og klassene som er definert av ulike spesialister.

Som et resultat av denne inkonsekvensen kan driftsdokumentasjonen som innkjøpsavdelingen (ERP) mottar når en byggevarer kjøpes inn fra en produsent, ofte ikke knyttes korrekt til klassifiseringen av varen på byggeplassen (PMIS, SCM). Som en konsekvens av dette er det lite sannsynlig at kritisk informasjon blir integrert i infrastruktur- og kapitalforvaltningssystemer (CAFM, AMS), noe som skaper alvorlige problemer under idriftsettelse, samt under påfølgende vedlikehold (AMS, RPM) eller utskifting av elementet.

I bedrifter med høy digital modenhet spiller klassifikatorer rollen som et nervesystem som forener all informasjonsflyt. Det samme elementet får en unik identifikator, slik at det kan overføres mellom CAD, ERP, AMS og CAFM-systemene og klassifikatorene deres uten forvrengning eller tap.

For å bygge effektive klassifikatorer må du forstå hvordan dataene brukes. Den sammeingeniøren kan navngi og klassifisere et element forskjellig i ulike prosjekter. Bare ved å samle inn bruksstatistikk over flere år kan man utvikle et stabilt klassifiseringssystem. Maskinlæring hjelper til med dette: Algoritmer analyserer tusenviser av prosjekter (fig. 9.1-10) og identifiserer sannsynlige klasser og parametere ved hjelp av maskinlæring (fig. 10.1-6). Automatisk klassifisering er spesielt verdifullt i miljøer der manuell klassifisering ikke er mulig på grunn av datamengden. Automatiske klassifiseringssystemer vil kunne skille mellom grunnleggende kategorier basert på minimalt med vareparametere (flere detaljer i bokens niende og tiende

del).

Utviklede klassifiseringssystemer blir katalysatorer for ytterligere digitalisering, og skaper grunnlag for

- Automatisert estimering av prosjektkostnader og tidslinjer.
- Forutseende analyse av potensielle risikoer og konflikter
- Optimalisering av innkjøpsprosesser og logistikkjeder
- Skape digitale tvillinger av bygninger og strukturer
- Integrasjoner med smartby- og tingenes internett-systemer

Tiden for transformasjon er begrenset - med utviklingen av maskinlæring og datasynsteknologier vil problemet med automatisk klassifisering, som har vært uløselig i flere tiår, bli løst i løpet av de kommende årene, og bygg- og designfirmaer som ikke klarer å tilpasse seg i tide, risikerer å gjenta skjebnen til drosjeflåter som blir fortrengt av digitale plattformer.

Automatisering av kostnads- og tidsplanberegninger, samt stordata og maskinlæring, vil bli diskutert mer inngående i bokens femte og niende del. Risikoen for en gjentakelse av drosjeflåtenes skjebne og Uberiseringen av byggebransjen diskuteres inngående i bokens tiende del.

For å forstå klassifiseringsverktøyenes nøkkelrolle i den digitale transformasjonen av byggebransjen, er det nødvendig å se nærmere på deres utviklingshistorie. Det er den historiske konteksten som gjør det mulig for oss å forstå hvordan tilnærmingen til klassifisering har utviklet seg, og hvilke trender som definerer dagens tilstand.

Masterformat, OmniClass, Uniclass og CoClass: utviklingen av klassifikasjonssystemer

Historisk sett har byggelement- og arbeidsklassifiseringsverktøyene utviklet seg i tre generasjoner, og hver av dem gjenspeiler nivået på tilgjengelig teknologi og de aktuelle behovene i bransjen i en bestemt tidsperiode (figur 4.2-8):

- **Første generasjon** (fra begynnelsen av 1950-tallet til slutten av 1980-tallet) - papirkataloger, hierarkiske klassifikatorer som brukes lokalt (f.eks. Masterformat, SfB).
- **Den andre generasjonen** (slutten av 1990-tallet til midten av 2010-tallet) er regneark og strukturerte databaser implementert i Excel og Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Tredje generasjon** (2010-tallet til i dag) - digitale tjenester og API-er -grensesnitt, integrasjon med CAD (BIM), automatisering (Uniclass 2015, CoClass).

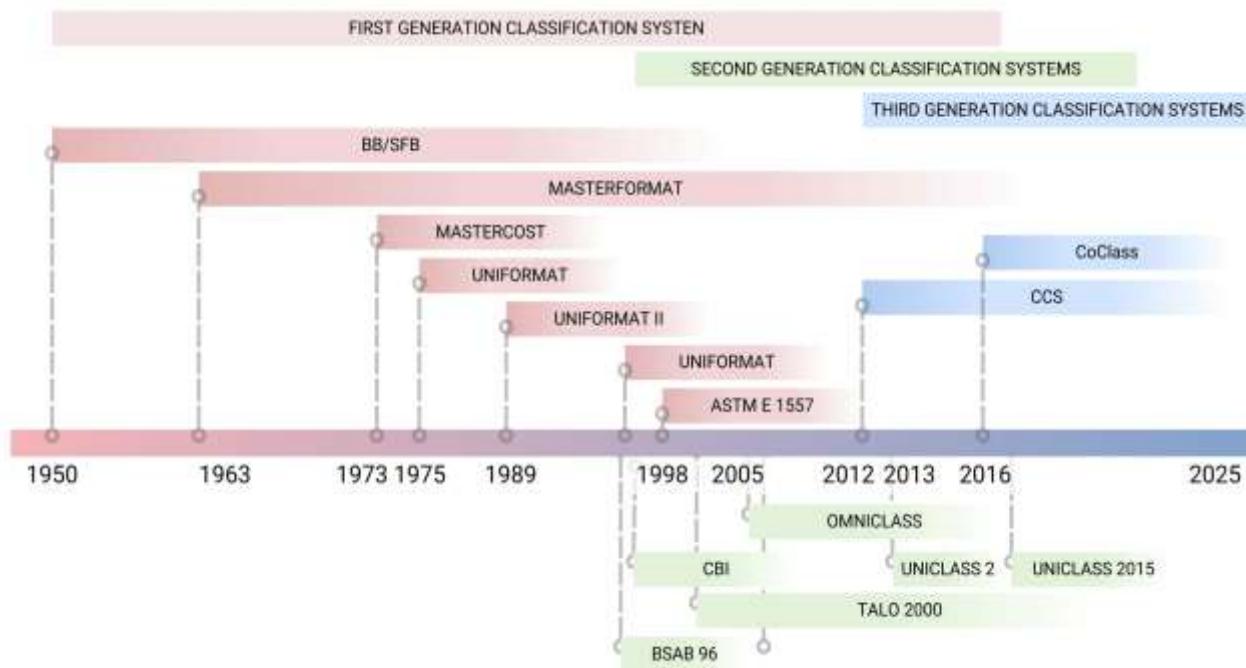


Fig. 4.2-8 Tre generasjoner av klassifikatorer for byggebransjen.

I løpet av de siste tiårene har den hierarkiske kompleksiteten (fig. 4.2-9) i klassifikatorer blitt redusert: Mens tidlige systemer som OmniClass, brukte opptil 7 nivåer for å beskrive 6887 klasser, er moderne løsninger som CoClass begrenset til 3 nivåer med 750 klasser. Dette gjør det enklere å arbeide med data, samtidig som den nødvendige granulariteten opprettholdes. Uniclass 2015, som ofte brukes som standard i Storbritannia, kombinerer 7210 klasser i bare 4 nivåer, noe som gjør den praktisk for CAD prosjekter og offentlige anskaffelser.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Fig. 4.2-9 For hver ny generasjon klassifiseringsverktøy reduseres kompleksiteten i kategoriseringen dramatisk.

På grunn av forskjeller i klassifisering kan selv et typisk element som en grunnmur av betong beskrives på svært forskjellige måter i de ulike landenes kalkulasjonssystemer (figur 4.2-10). Disse forskjellene gjenspeiler nasjonal bygeskikk, målesystemer som brukes, tilnærminger til materialklassifisering og de regulatoriske og tekniske kravene som gjelder i hvert land.



Fig. 4.2-10 Det samme elementet brukes i prosjekter i forskjellige land gjennom ulike beskrivelser og klassifiseringer.

De mange ulike klassifiseringene av de samme elementene kompliserer det internasjonale samarbeidet og gjør det tidkrevende og noen ganger nesten umulig å sammenligne kostnader og arbeidsomfang i internasjonale prosjekter. For øyeblikket finnes det ikke én universell klassifisering på globalt nivå - hvert land eller hver region utvikler sine egne systemer basert på lokale normer, språk og forretningskultur:

- **CCS** (Danmark): Cost Classification System - et system for klassifisering av kostnader gjennom hele livssyklusen til et anlegg (design, bygging, drift). Hovedvekten ligger på drifts- og vedlikeholdslogikk, men inkluderer også budsjett- og ressursstyring.
- **NS 3451** (Norge): Kategoriserer anlegg etter funksjon, designelementer og livssyklusfaser. Brukes til prosjektstyring, kostnadsestimering og langsiktig planlegging.
- **MasterFormat** (USA): et system for strukturering av byggespesifikasjoner i seksjoner (f.eks. betong, elektro, etterbehandling). Fokus på fagområder og arbeidstyper i stedet for funksjonelle elementer (i motsetning til UniFormat).
- **Uniclass 2** (UK): en av de mest detaljerte klassifikatorene, som brukes i offentlige anskaffelser og BIM prosjekter. Samler data om objekter, arbeider, materialer og rom i ett enkelt system.
- **OmniClass**: en internasjonal standard (utviklet av CSI i USA) for håndtering av objektinformasjon fra komponentbiblioteker til elektroniske spesifikasjoner. Egnet for langtidslagring av data, kompatibel med CAD (BIM) og andre digitale verktøy.
- **COBie**: Construction-Operation Building information exchange er en internasjonal standard for utveksling av data mellom prosjekterings-, bygge- og driftsfasen. Inkludert i BS 1192-4:2014 som en del av konseptet "BIM - modell klar til bruk". Fokuserer på informasjonsoverføring (f.eks. utstyrsspesifikasjoner, garantier, entreprenørkontakter).

Globaliseringen av byggebransjen vil sannsynligvis føre til en gradvis harmonisering av klassifikasjonssystemene for bygningselementer, noe som vil redusere avhengigheten av lokale, nasjonale standarder betraktelig. Denne prosessen kan utvikle seg i tråd med utviklingen av Internett-kommunikasjon, der universelle dataoverføringsprotokoller etter hvert erstattet ulike lokale formater og sikret global interoperabilitet mellom systemer.

En alternativ utviklingsvei kan være en direkte overgang til automatiske klassifiseringssystemer basert på maskinlæringsteknologi. Disse teknologiene, som i dag hovedsakelig utvikles innen autonom transport, har et betydelig potensial for anvendelse på store CAD-designdatasett (fig. 10.1-6).

I dag er situasjonen ikke bare begrenset til den nasjonale klyngen av klassifiseringsenheter. På grunn av de mange særegenheter som ikke er tatt hensyn til på nasjonalt nivå, må hvert enkelt selskap forene og standardisere kategoriene av elementer og ressurser som det arbeider med.

Som regel starter denne prosessen i det små - med lokale tabeller over objekter eller interne merkesystemer. Det strategiske målet er imidlertid å gå over til et felles språk for å beskrive alle elementer, som er forståelig ikke bare innad i bedriften, men også utenfor - ideelt sett harmonisert med internasjonale klassifikatorer eller bransjeklassifikatorer (fig. 4.2-8). Denne tilnærmingen gjør det lettere å integrere med eksterne partnere og digitale systemer, og bidrar til å skape enhetlige ende-til-ende-prosesser i objektenes livssyklus.

Før man går over til automatisering og skalerbare IT-systemer, må man enten bruke klassifikatorer på nasjonalt nivå eller bygge opp sin egen logiske og entydige struktur for identifikasjon av elementer. Hvert objekt - det være seg et vindu (figur 4.2-11), en dør eller et teknisk system - må beskrives på en slik måte at det kan gjenkjennes på en entydig måte i alle virksomhetens digitale systemer. Dette er avgjørende i overgangen fra flate tegninger til digitale modeller, både i prosjekteringsfasen og i driften av bygningen.

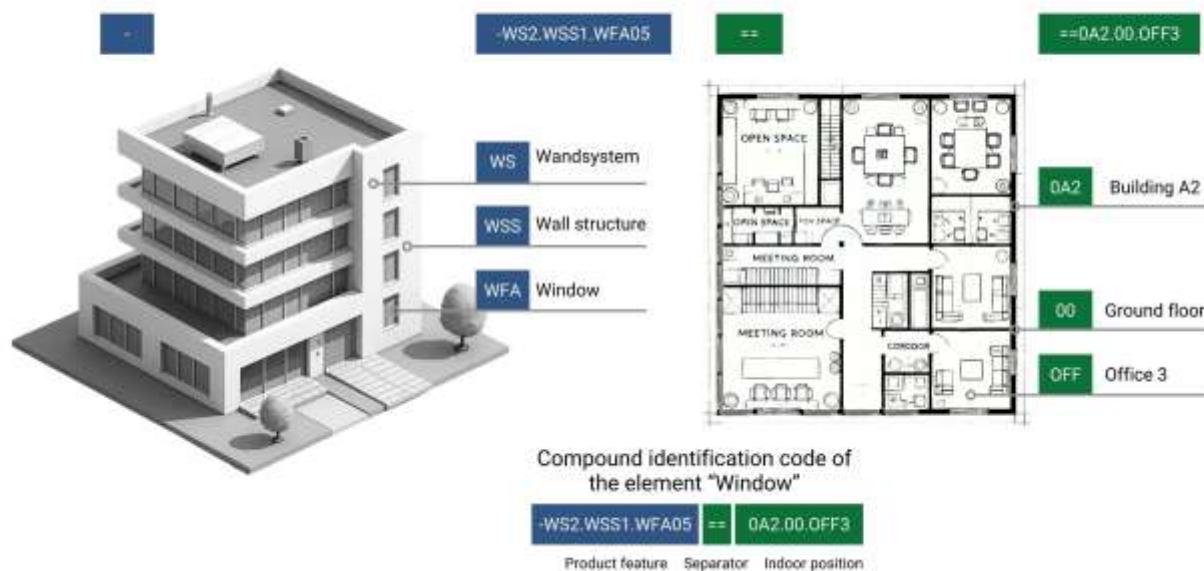


Fig. 4.2-11 Eksempel på sammensatt vindusbygningselementidentifikator basert på klassifisering og plassering i bygningen.

Et eksempel på interne klassifikatorer kan være utviklingen av en sammensatt identifikasjonskode (figur 4.2-11). En slik kode kombinerer flere nivåer av informasjon: elementets funksjonelle formål (f.eks. "vindu i vegg"), dets type og den nøyaktige romlige referansen - bygning A2, etasje 0, rom 3. En slik flernivåstruktur gjør det mulig å skape et enhetlig system for navigering gjennom digitale modeller og dokumentasjon, spesielt i stadier av dataverifikasiing og -transformasjon, der det er nødvendig med entydig gruppering av elementer. Entydig elementgjenkjenning sikrer konsistens mellom avdelingene og reduserer risikoen for duplisering, feil og tap av informasjon.

En velutviklet klassifikator er ikke bare et teknisk dokument, det er selve fundamentet i et selskaps digitale økosystem:

- sikrer datakompatibilitet mellom systemer;
- reduserer kostnadene ved å søke etter og behandle informasjon;
- øker gjennomsiktigheten og håndterbarheten;
- skaper et grunnlag for skalering og automatisering.

Den standardiserte beskrivelsen av objekter, gjennom bruk av nasjonale klassifikatorer eller proprietære sammensatte identifikasjonskoder, blir grunnlaget for konsistente data, pålitelig informasjonsutveksling og påfølgende implementering av intelligente tjenester - fra automatiserte innkjøp til digitale tvillinger.

Etter å ha strukturert data i flere formater og valgt klassifikatoren som skal brukes til å gjenkjenne og gruppere elementene, er neste trinn å modellere dataene riktig. Denne prosessen innebærer å identifisere nøkkelparametere, bygge opp en logisk datastruktur og beskrive relasjonene mellom elementene.



KAPITTEL 4.3.

DATAMODELLERING OG KOMPETANSESENTER

Datamodellering: konseptuell, logisk og fysisk modell

Effektiv håndtering av data (som vi tidligere har strukturert og kategorisert) er umulig uten en gjennomtenkt lagrings- og behandlingsstruktur. For å sikre tilgang til og konsistens i informasjonen i lagrings- og prosesseringsfasen bruker bedrifter datamodellering, en metodikk som gjør det mulig å utforme tabeller, databaser og koblinger mellom dem i samsvar med virksomhetens krav.

Datamodellering er fundamentet som ethvert digitalt økosystem bygger på. Uten en beskrivelse av systemer, krav og datamodellering vet eller forstår ikke ingeniører og fagfolk som skaper data, hva dataene de skaper, skal brukes til.

På samme måte som man ikke kan begynne å legge murstein uten en plan når man skal bygge en bygning, må man ha en klar forståelse av hvilke data som skal brukes, hvordan de skal kobles sammen, og hvem som skal jobbe med dem når man skal lage et datavarehus. Uten en beskrivelse av prosesser og krav mister ingeniørene og fagfolkene som skaper dataene, oversikten over hvor og hvordan dataene skal brukes i fremtiden.

Datamodellen fungerer som en bro mellom virksomheten og IT. Den gjør det mulig å formalisere krav, strukturere informasjon og lette kommunikasjonen mellom interessentene. Slik sett kan datamodellering sammenlignes med arbeidet til en arkitekt som, i henhold til kundens plan, utvikler en byggeplan og deretter sender den videre til byggherrene - databaseadministratører og utviklere - for implementering (opprettelse av databasen).

I tillegg til å strukturere og kategorisere elementer og ressurser (fig. 4.2-11), må alle byggfirmaer derfor beherske kunsten å "bygge" databaser (tabeller) og lære seg å skape koblinger mellom dem, som om de kobler sammen mursteinene til en pålitelig og sterk vegg av kunnskap fra bedriftens data. Viktige begreper i datamodellering (figur 4.3-1) er blant annet

- **Enheter** er objekter som det må samles inn data om. I den tidlige prosjekteringsfasen kan en enhet være et enkelt element (f.eks. "dør"), og i estimeringsmodellen kan det være en gruppe elementer som er kategorisert (f.eks. "innvendige dører").
- **Attributter** er egenskaper ved enheter som beskriver viktige detaljer: dimensjoner, egenskaper, monteringskostnader, logistikk og andre parametere.
- **Relasjoner (koblinger)** - viser hvordan enheter interagerer med hverandre. De kan være av en av følgende typer: "en til en", "mange til en", "mange til mange".
- ER-diagrammer (Entity-Relationship-diagrammer) er visuelle diagrammer som viser entiteter, attributter og relasjonene mellom dem. ER-diagrammer kan være konseptuelle, logiske og fysiske - hver av dem gjenspeiler et annet detaljnivå.

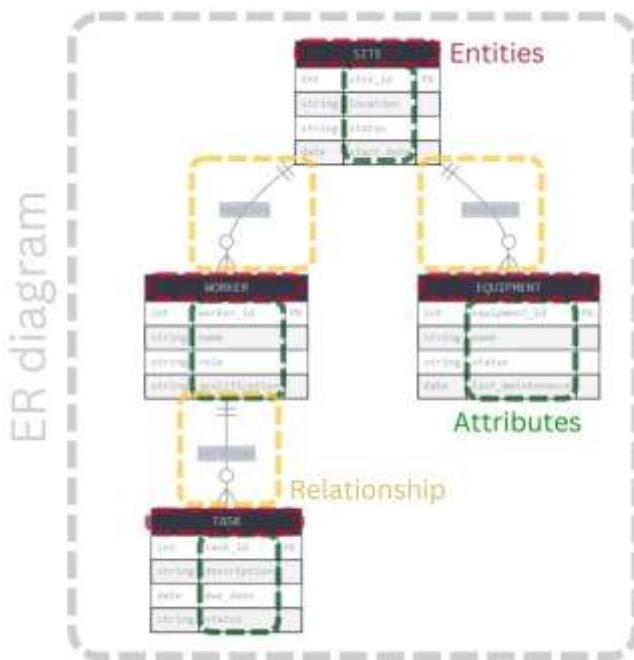


Fig. 4.3-1 ER-diagram over en konseptuell databasestruktur med entiteter, attributter og relasjoner.

Prosesssen med å designe data og definere relasjoner mellom dem deles tradisjonelt inn i tre hovedmodeller. Hver av dem oppfyller visse funksjoner, og de varierer i detaljnivå og abstraksjonsgrad når det gjelder representasjon av datastrukturen:

- **Konseptuell datamodell:** Denne modellen beskriver de viktigste enhetene og relasjonene mellom dem uten å gå inn på attributt-detaljer. Den brukes vanligvis i de innledende fasene av planleggingen. På dette stadiet kan vi skissere fra databaser og systemer for å vise forholdet mellom ulike avdelinger og spesialister.

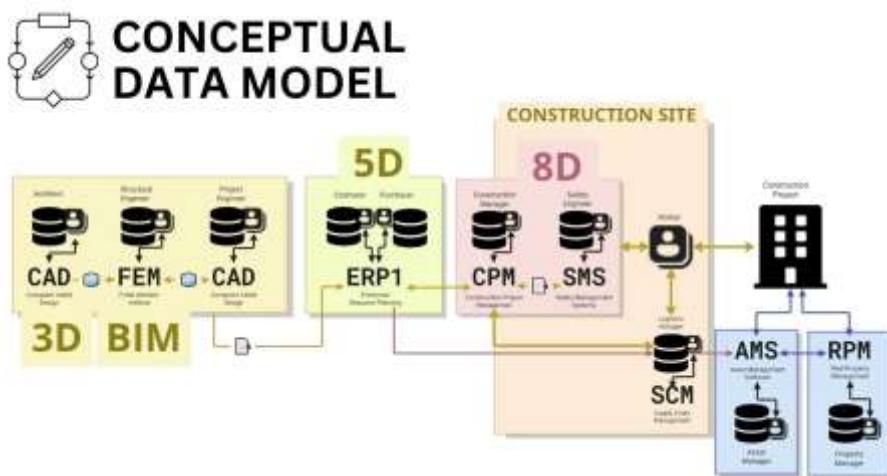
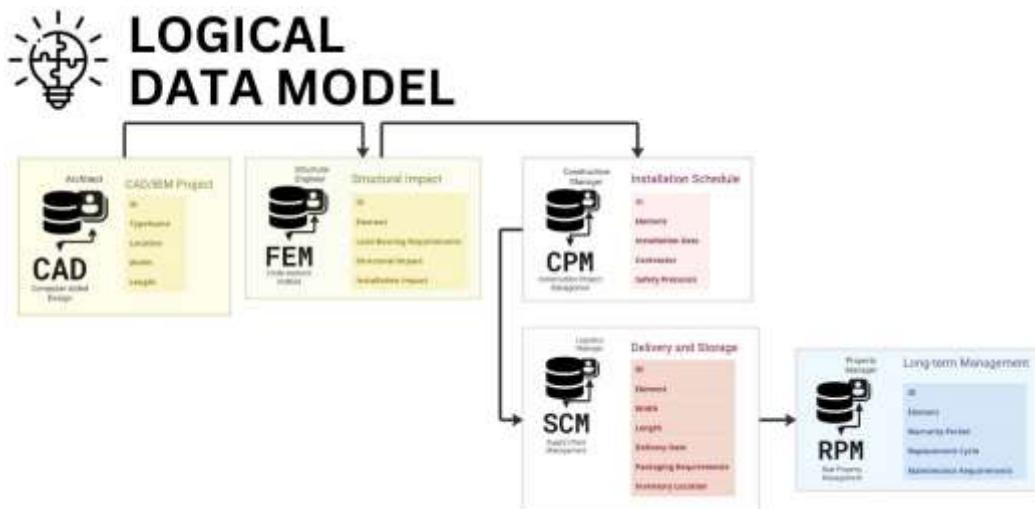


Fig. 4.3-2 Konseptdiagrammet beskriver innholdet i systemet: en fremstilling på høyt nivå av sammenhengene, uten tekniske detaljer.

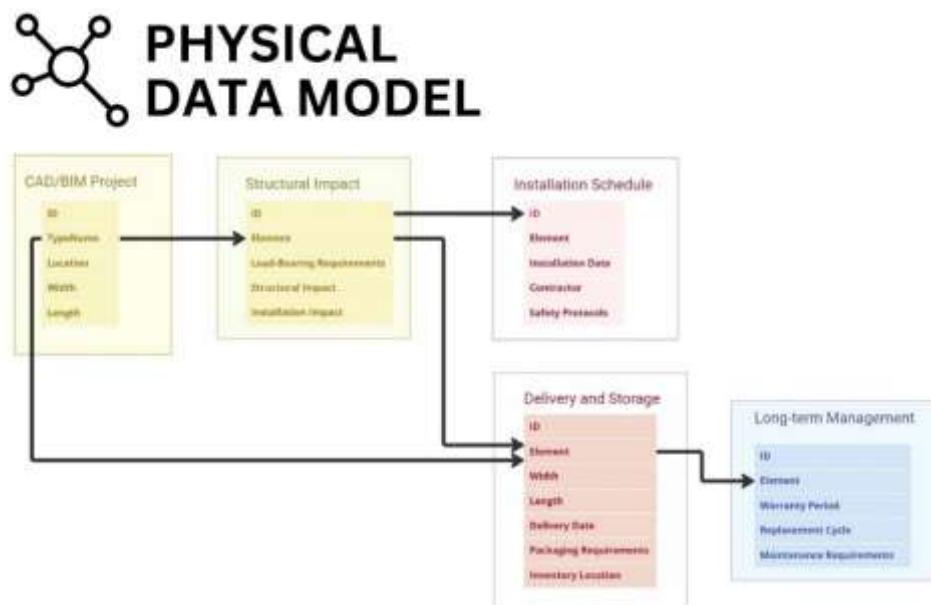
- **Logisk datamodell:** Basert på den konseptuelle modellen inneholder den logiske datamodellen

detaljerte beskrivelser av enheter, attributter, nøkler og relasjoner, og kartlegger forretningsinformasjon og -regler.



Figur 4.3-3 Logisk datamodell beskriver datatyper, relasjoner og nøkler i detalj, men uten systemimplementering.

- **Fysisk datamodell:** Denne modellen beskriver de nødvendige strukturene for å implementere en database, inkludert tabeller, kolonner og relasjoner. Den fokuserer på databaseytelse, indeksningsstrategier og fysisk lagring for å optimalisere den fysiske distribusjonen av databaser.



Figur 4.3-4 Den fysiske datamodellen definerer hvordan systemet skal implementeres, inkludert tabeller og spesifikke detaljer i databasen.

Når man utformer databaser og tabellrelasjoner, er det viktig å forstå abstraksjonsnivåene for å kunne bygge en effektiv systemarkitektur.

En effektiv datamodelleringsmetodikk gjør det mulig å kombinere forretningsmessige mål med teknisk implementering, noe som gjør hele prosesskjeden mer oversiktlig og håndterbar. Datamodellering er ikke en engangsoppgave, men en prosess som består av sekvensielle trinn (figur 4.3-5):

- **Innhenting av forretningskrav:** Nøkkelloppgaver, mål og informasjonsflyt defineres. Dette er fasen der man samhandler aktivt med eksperter og brukere.
- **Enhetsidentifikasjon:** De viktigste objektene, kategoriene og datatypene som er viktige å ta hensyn til i det fremtidige systemet, fremheves.
- **Utvikling av en konseptuell og logisk modell:** Først kartlegges de viktigste enhetene og relasjonene mellom dem, deretter attributter, regler og detaljert struktur.
- **Fysisk modellering:** Den tekniske implementeringen av modellen utformes: tabeller, felt, relasjoner, begrensninger, indeks.
- **Opprettelse av database:** Det siste trinnet er å implementere den fysiske modellen i det valgte DBMS, utføre testing og klargjøre for drift.

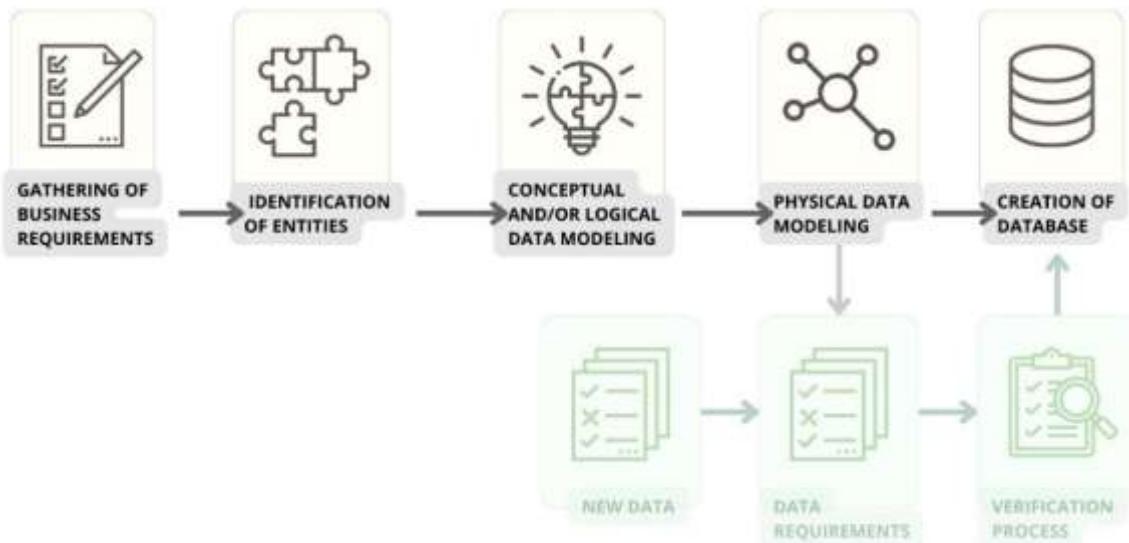


Fig. 4.3-5 Opprettelse av databaser og datahåndteringssystemer for forretningsprosesser begynner med generering av krav og datamodellering.

Riktig utformede datamodelleringsprosesser gir oversiktlig informasjonsflyt, noe som er spesielt viktig i komplekse prosjekter som byggeprosjekter eller anleggsledelse. La oss se på hvordan det å gå fra en konseptuell modell til en logisk modell og deretter til en fysisk modell kan bidra til å effektivisere prosessene.

Praktisk datamodellering i forbindelse med bygg- og anleggsvirksomhet

La oss ta en byggeplassledelsesoppgave som et eksempel på datamodellering og konvertere kravene til formann til en strukturert logisk modell. Med utgangspunkt i de grunnleggende behovene for byggeplassledelse definerer vi nøkkelenheter for: byggeplass (SITE), arbeidere (WORKER), utstyr (EQUIPMENT), oppgaver (TASK) og utstyrss bruk (EQUIPMENT_USAGE). Hver entitet inneholder et sett med attributter som gjenspeiler viktige egenskaper. For TASK kan dette for eksempel være en beskrivelse av oppgaven, forfallsdato, status, prioritet; for WORKER kan det være navn, rolle på anlegget, nåværende ansettelse osv.

Logikkmodellen etablerer relasjoner mellom disse enhetene og viser hvordan de samhandler med hverandre i reelle arbeidsprosesser (figur 4.3-6). For eksempel viser forholdet mellom anlegg og arbeidere at mange arbeidere kan arbeide på ett anlegg, mens forholdet mellom arbeidere og oppgaver gjenspeiler at én arbeider kan utføre flere oppgaver.

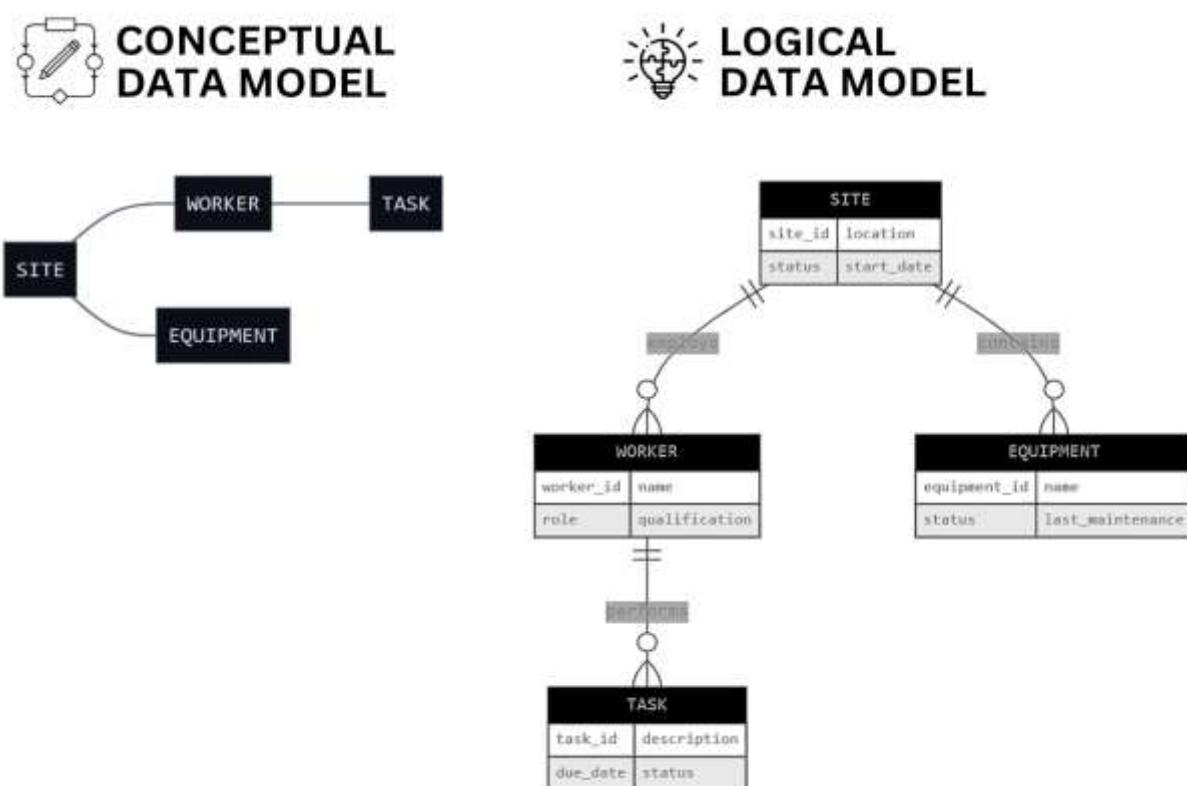
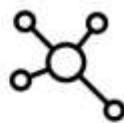


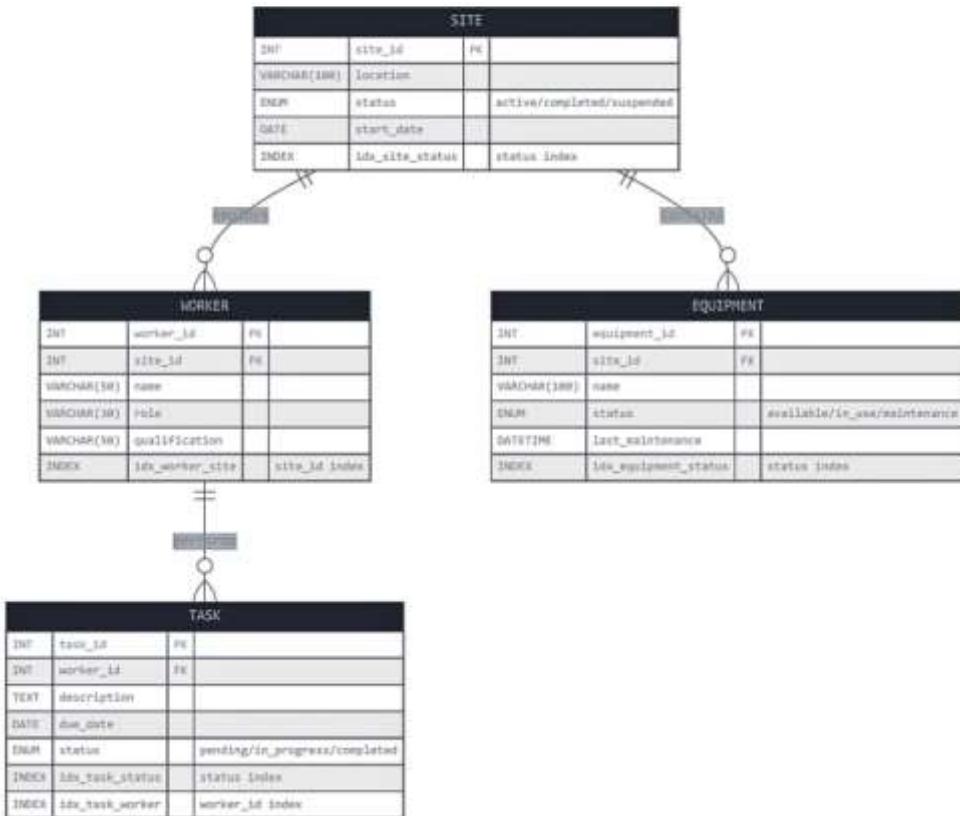
Fig. 4.3-6 Konseptuell og logisk datamodell generert av formannens krav til beskrivelse av byggeplassens prosesser.

Når man går over til den fysiske modellen, legger man til tekniske implementeringsdetaljer: spesifikke datatyper (VARCHAR, INT, DATE), primær- og fremmednøkler for relasjoner mellom tabeller og indeks for å optimalisere databasens ytelse (figur 4.3-7).

For eksempel bør det defineres spesifikke typer med mulige verdier for statuser, og indeks på nøkkelfelt som status og worker_id bør legges til for å forbedre søkeytelsen. På denne måten blir den logiske beskrivelsen av systemet til en konkret implementeringsplan for databasen, klar til å opprettes og implementeres.



PHYSICAL DATA MODEL



Figur 4.3-7 Den fysiske datamodellen beskriver enhetene på en byggeplass ved hjelp av de minste parameterne som kreves.

Den fysiske modellen er ofte forskjellig fra den logiske modellen. I gjennomsnitt er fordelingen av modelleringstiden som følger: ca. 50% går med til den konseptuelle modellen (innsamling av krav, diskusjon av prosesser, identifisering av enheter), 10% til den logiske modellen (spesifisering av attributter og relasjoner) og 40% til den fysiske modellen (implementering, testing, tilpasning til DBMS).

Denne balansen skyldes at den konseptuelle fasen legger grunnlaget for datastrukturen, mens den logiske modellen kun spesifiserer relasjoner og attributter. Den fysiske modellen krever mest ressurser, ettersom det er på dette stadiet dataene implementeres i spesifikke plattformer og verktøy.

Opprette en database ved hjelp av LLM

Når vi har en datamodell og beskrivelse av enheter gjennom parametere, er vi klare til å opprette databaser - lagre, der vi vil lagre informasjon som kommer etter struktureringfasen på spesifikke prosesser.

La oss prøve å lage et eksempel på en enkel, men funksjonell database med et minimum av kode ved hjelp av SQLite ved hjelp av eksemplet med programmeringsspråket Python. Relasjonsdatabaser ble diskutert i detalj i kapittelet "Strukturerte relasjonsdatabaser og spørrespråket SQL".

SQLite er en lett, åpen kildekodebasert, innebygd database som kan brukes på flere plattformer, og som støtter sentrale SQL-funksjoner. SQLite er mye brukt i CAD (BIM), ERP, EPM og finnes i nesten alle smarttelefoner.

La oss be LLM -modellen om å opprette en liten SQLite-database for oss basert på den fysiske modellen som ble opprettet ovenfor (figur 4.3-4) for FEM-systemet (strukturell påvirkning) og CPM (monteringsplan).

- Send en tekstforespørsel til LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller andre):

Opprett SQLite-databaser for systemene "Constructive Impact" og "Installation Schedule" fra figuren i vedlegget (lim inn skjemaet i chatten som et bilde (Fig. 4.3-4) ↴

2 LLMs svar :

```
1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11                 ID INTEGER PRIMARY KEY,
12                 Element TEXT,
13                 LoadBearingRequirements TEXT,
14                 StructuralImpact TEXT,
15                 InstallationImpact TEXT
16             )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20                 ID INTEGER PRIMARY KEY,
21                 Element TEXT,
22                 InstallationDate DATE,
23                 Contractor TEXT,
24                 SafetyProtocols TEXT
25             )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")
```

Run in IDE

Figur 4.3-8 Den genererte koden oppretter databaser og tabeller fra det opplastede bildet, som vi nå kan laste inn data i.

Denne databasekoden kan kjøres i Python-miljøet med SQLite-støtte, som vanligvis er inkludert i standard Python-miljø. Den kan kjøres på en hvilken som helst lokal Python-maskin som er i stand til å kjøre skript og opprette filer. Når koden er kjørt og utført, vil den opprette en fil på harddisken din - en SQLite-database med navnet `construction.db` som ligner på en Excel-fil med regneark) som inneholder tabellene "Structural Impact" og "Installation Schedule".

Når en ny **construction.db-database** er opprettet i disse tabellene, kan data legges til i disse tabellene via SQL _queries eller import, slik at du senere kan begynne å lage automatisk databehandling. Data kan importeres til SQLite-databasen fra CSV-filer, Excel-regneark eller eksporteres fra andre databaser og repositorier via API.

For å etablere en bærekraftig datamodellering og effektive prosesser for databaseadministrasjon trenger en bedrift en klart definert strategi og koordinering mellom tekniske og forretningsmessige team. Med ulike prosjekter og flere datakilder er det ofte vanskelig å sikre konsistens, standardisering og kvalitetskontroll på alle nivåer. En viktig løsning kan være å opprette et eget kompetansesenter for datamodellering (CoE) i selskapet.

Senter for fremragende forskning (CoE) for datamodellering

Etter hvert som data blir en av de viktigste strategiske ressursene, må bedriftene gjøre mer enn bare å samle inn og lagre informasjon på riktig måte - det er viktig å lære seg å håndtere data systematisk. Centre of Excellence for Classification and Data Modelling (CoE) er en strukturell enhet som sikrer konsistens, kvalitet og effektivitet i all datahåndtering i organisasjonen.

Centre of Excellence (CoE) er kjernen i ekspertstøtten og det metodiske grunnlaget for digital transformasjon i en bedrift. Det bygger en datadrevet kultur og gjør det mulig for organisasjoner å bygge prosesser som tar beslutninger basert på strukturerte, validerte og representative data i stedet for på intuisjon eller lokal informasjon.

Et datasenter for ekspertise består vanligvis av tverrfunksjonelle team som arbeider etter "to pizzaer"-prinsippet. Dette prinsippet, som ble foreslått av Jeff Bezos, innebærer at størrelsen på teamet skal være slik at det kan mates med to pizzaer, dvs. ikke mer enn 6-10 personer. Denne tilnærmingen bidrar til å unngå overdreven byråkratisering og øker fleksibiliteten i arbeidet. CoE-teamet bør bestå av medarbeidere med ulike tekniske ferdigheter, fra dataanalyse og maskinlæring til ekspertise innen spesifikke forretningsområder. Med sin dype tekniske kunnskap skal dataingeniørene ikke bare optimalisere prosesser og modellere data, men også støtte kolleger ved å redusere tiden som brukes på rutineoppgaver (fig. 4.3-9).

På samme måte som økosystemenes robusthet i naturen sikres gjennom biologisk mangfold, oppnås fleksibilitet og tilpasningsevne i den digitale verden gjennom et mangfold av tilnærminger til datahåndtering. Dette mangfoldet må imidlertid underbygges av felles regler og konsepter.

Centre of Excellence (CoE) kan sammenlignes med "klimaforholdene" i et skogøkosystem, som avgjør hvilke typer data som vil blomstre og hvilke som automatisk vil bli forkastet. Ved å skape et gunstig "klima"

for kvalitetsdata legger kompetansesenteret til rette for et naturlig utvalg av beste praksis og metoder, som deretter blir standarder for organisasjonen.

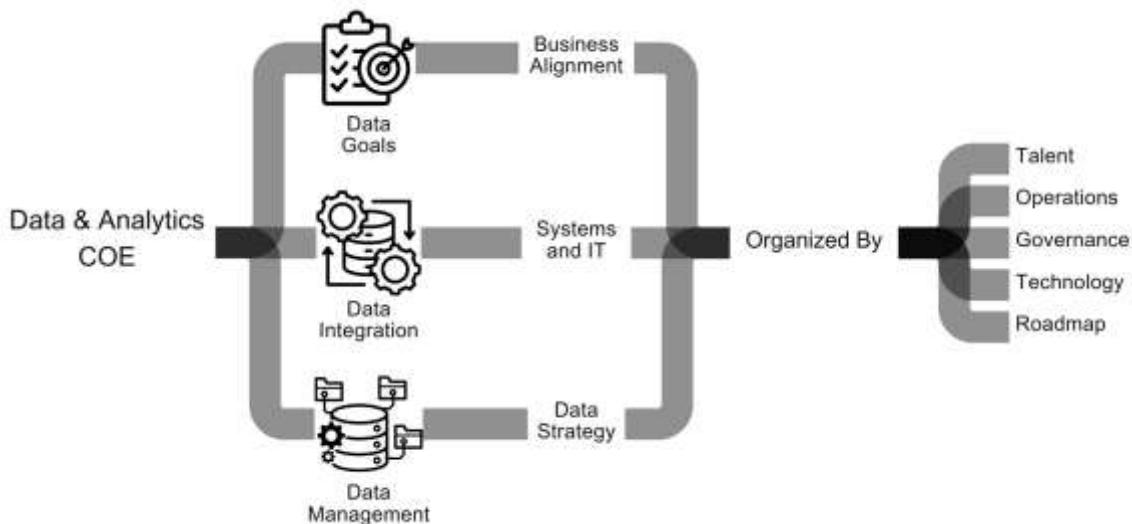


Fig. 4.3-9 Kompetansesenteret (CoE) for data og analyse samler ekspertise på viktige aspekter ved datahåndtering, -integrasjon og -strategi.

For å få raskere integrasjonssykluser og oppnå bedre resultater bør CoE gi medlemmene en tilstrekkelig grad av selvstendighet i beslutningsprosessen. Dette er spesielt viktig i et dynamisk miljø der prøving og feiling, stadige tilbakemeldinger og hyppige lanseringer kan gi betydelige fordeler. Denne autonomien er imidlertid bare effektiv hvis det er tydelig kommunikasjon og støtte fra toppledelsen. Uten en strategisk visjon og koordinering på toppnivå kan selv det mest kompetente teamet støte på hindringer når de skal gjennomføre sine initiativer.

Det er COE eller toppledelsen i selskapet som er ansvarlig for å sikre at datamodelleringstilnærmingen ikke er begrenset til ett eller to prosjekter, men er integrert i det overordnede systemet for informasjonsstyring og styring av forretningsprosesser.

I tillegg til oppgaver knyttet til datamodellering og datastyring er kompetansesenteret (CoE) ansvarlig for utviklingen av felles standarder og tilnærmingen til implementering og drift av datainfrastrukturen. I tillegg fremmer det en kultur for kontinuerlig forbedring, prosessoptimalisering og effektiv bruk av data i organisasjonen (figur 4.3-10).

Den systematiske tilnærmingen til data- og modellforvaltning i CoE kan grovt sett deles inn i flere hovedblokker:

- **Standardisering av prosesser og styring av modellenes livssyklus:** CoE utvikler og implementerer metoder for å samordne opprettelsen og styringen av datamodeller. Dette omfatter blant annet etablering av strukturmaler, kvalitetstilkontrollmetoder og systemer for versjonskontroll for å sikre datakontinuitet i alle faser av arbeidet.
- **Rollehåndtering og ansvarsfordeling:** COE definerer nøkkelrollene i datamodellingsprosessen. Alle prosjektdeltakerne tildeles klart definerte roller og ansvarsområder, noe som letter teamarbeidet og reduserer risikoen for uoverensstemmelser i dataene.

- **Kvalitetskontroll og revisjon:** Effektiv forvaltning av byggdata krever kontinuerlig overvåking av kvaliteten. Automatiserte mekanismer for kontroll av data, identifisering av feil og manglende attributter er under implementering.
- **Forvaltning av metadata og informasjonsarkitektur:** CoE er ansvarlig for å skape et enhetlig system for klassifisering og identifikatorer, standarder for navngivning og entitetsbeskrivelser, noe som er avgjørende for integrasjon mellom systemer.

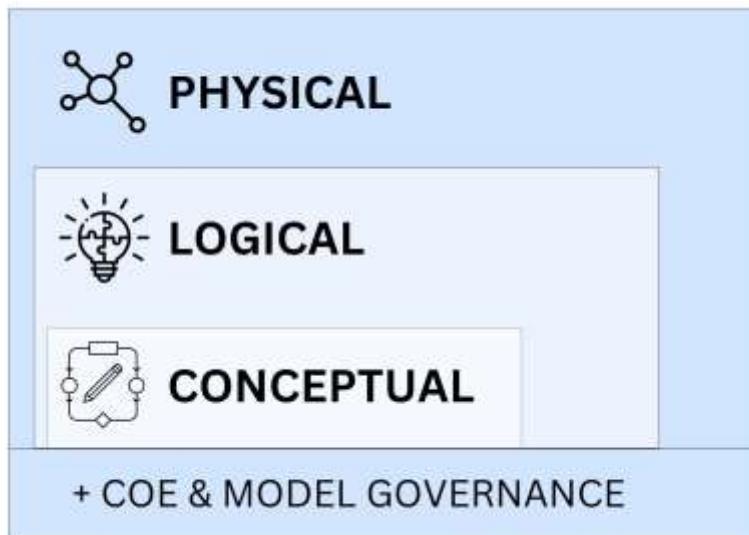


Fig. 4.3-10 Datamodellering og datakvalitetsstyring er en av hovedutfordringene for CoE

Centre of Excellence (CoE) for Data er ikke bare en gruppe eksperter, men en systemisk mekanisme som skaper en ny datadrevet kultur og sikrer en enhetlig tilnærming til arbeidet med data i hele selskapet. Gjennom kompetent integrering av modellingsprosesser i det overordnede informasjonsstyringssystemet, standardisering, klassifisering og datakvalitetskontroll, hjelper CoE virksomheter med å kontinuerlig forbedre produktene og forretningsprosessene sine, reagere raskere på markedsendringer og ta velbegrunnede beslutninger basert på pålitelige analyser.

Slike sentre er spesielt effektive når de kombineres med moderne DataOps-prinsipper - under et grep som sikrer kontinuerlig levering, automatisering og kvalitetskontroll av data. Vi kommer tilbake til DataOps i del 8, i kapittelet "Moderne data teknologi i byggebransjen".

I de følgende kapitlene går vi fra strategi til praksis - vi "transformerer" oss betinget til et datasenter: Vi skal se på flere eksempler på hvordan oppgaveparametrising, kravinnsamling og den automatiske valideringsprosessen foregår.



KAPITTEL 4.4.

SYSTEMATISERING AV KRAV OG VALIDERING AV INFORMASJON

Innhenting og analyse av krav: omforming av kommunikasjon til strukturerte data

Innsamling og håndtering av krav er det første skrittet for å sikre datakvalitet. Til tross for utviklingen av digitale verktøy formuleres de fleste krav fortsatt i ustukturert form: gjennom brev, møtereferater, telefonsamtaler og muntlige diskusjoner. Denne formen for kommunikasjon gjør det vanskelig å automatisere, validere og gjenbruke informasjon. I dette kapittelet ser vi på hvordan vi kan oversette tekstlige krav til formelle strukturer, slik at forretningskravene blir transparente og systematiske.

Gartners undersøkelse "Data Quality: Best Practices for Accurate Insights" fremhever den kritiske betydningen av datakvalitet for vellykkede data- og analyseinitiativer [100]. De påpeker at dårlig datakvalitet i gjennomsnitt koster organisasjoner minst 12,9 millioner dollar årlig, og at pålitelige data av høy kvalitet er avgjørende for å skape et datadrevet selskap.

Mangelen på strukturerte krav fører til at det samme elementet (entiteten) og dets parametere kan være lagret i forskjellige systemer i ulike varianter. Dette reduserer ikke bare effektiviteten i prosessene, men fører også til bortkastet tid, duplisering av informasjon og behov for å validere data på nytt før de kan brukes. Resultatet er at selv en enkelt utelatelse - en tapt parameter eller et enkelt element som er feil beskrevet - kan forsinke beslutningsprosessen og føre til ineffektiv ressursbruk.

*I mangel av en spiker gikk hesteskoen tapt.
I mangel av en hesteskko gikk hesten tapt.
I mangel av en hest, gikk rytteren tapt.
I mangel av en ryttar, gikk et budskap tapt.
I mangel av et budskap gikk slaget tapt.
I mangel av et slag gikk et kongerike tapt.
Alt sammen fordi det manglet en spiker i hesteskoen.*

- Ordspråk [101]

Analyse og innsamling av krav til prosessen for utfylling og lagring av data starter med å identifisere alle interesser. På samme måte som tapet av en enkelt spiker fører til en kjede av kritiske konsekvenser, kan tapet av en enkelt interessent, et oversett krav eller til og med en enkelt parameter få store konsekvenser ikke bare for en enkelt forretningsprosess, men for hele økosystemet i et prosjekt og for organisasjonen som helhet. Derfor er det avgjørende å identifisere selv de elementene, parameterne og rollene som ved første øyekast virker ubetydelige, men som senere kan vise seg å være kritiske for virksomhetens bærekraft.

La oss tenke oss at en bedrift har et prosjekt der kunden kommer med en ny forespørsel - "legg til et ekstra vindu på nordsiden av bygningen". Den lille prosessen "kundens forespørsel om å legge til et nytt vindu i det nåværende prosjektet" involverer arkitekt, kunde, CAD-spesialist (BIM), byggeleder, logistikksjef, ERP - analytiker, kvalitetskontrollingeniør, sikkerhetsingeniør, kontrollsjef og eiendomsforvalter.

Selv en liten prosess kan involvere mange titalls forskjellige spesialister. Hver enkelt prosessdeltaker må forstå kravene til spesialistene de er koblet sammen med på datanivå.

På tekstnivå (fig. 4.4-1) foregår kommunikasjonen mellom kunden og spesialistene i prosesskjeden på følgende måte:

- ⦿ **Kunde:** "Vi har bestemt oss for å sette inn et ekstra vindu på nordsiden for å få bedre belysning. Kan dette realiseres?"
- ⦿ **Arkitekt:** "Ja visst, jeg skal revidere prosjektet for å inkludere det nye vinduet og sende oppdaterte CAD-tegninger til (BIM)".
- ⦿ **CAD-spesialist (BIM):** "Mottar et nytt prosjekt. Jeg oppdaterer CAD (BIM)-modellen med det ekstra vinduet, og etter koordinering med FEM-ingeniøren oppgir jeg den nøyaktige plasseringen og dimensjonene til det nye vinduet."
- ⦿ **Byggeleder:** "Vi har mottatt et nytt prosjekt. Vi justerer installasjonsdatoene til 4D og informerer alle relevante underleverandører."
- ⦿ **Anleggssingeniør (CAFM):** "Jeg vil legge inn dataene fra 6D i det nye vinduet i CAFM-systemet for fremtidig anleggsstyring og vedlikeholdsplanlegging."
- ⦿ **Logistikksjef:** "Jeg trenger målene og vekten på det nye vinduet for å organisere leveringen av vinduet til byggeplassen".
- ⦿ **ERP-analytiker:** "Jeg trenger omfangstabellene og den nøyaktige vindustypen for 5D budsjettoppdateringen i ERP-systemet vårt for å gjenspeile kostnadene for det nye vinduet i det samlede prosjektestimatet."
- ⦿ **Kvalitetskontrollingeniør:** "Når vindusspesifikasjonene er klare, skal jeg sørge for at de oppfyller våre kvalitets- og materialstandarder."
- ⦿ **Sikkerhetsingeniør:** "Jeg skal vurdere sikkerhetsaspektene ved det nye vinduet, med særlig fokus på samsvar og evakuering i henhold til 8D-ordningen".
- ⦿ **Kontrollsjef:** "Basert på det nøyaktige arbeidsomfanget fra ERP vil vi oppdatere vår 4D-tidslinje for å gjenspeile installasjonen av det nye vinduet, og lagre de nye dataene i prosjektets innholdsstyringssystem."
- ⦿ **Arbeider (installatør):** "Trenger instruksjoner om installasjon, montering og tidspunkt for arbeidet. Er det i tillegg innført noen spesielle sikkerhetsregler som jeg må følge?"
- ⦿ **Eiendomsforvalter:** "Når det er installert, vil jeg dokumentere garanti- og vedlikeholdsinformasjon for langsigktig forvaltning".
- ⦿ **Asset Manager:** "Utstyrsingeniør, vennligst send endelige data for sporing og livssyklusstyring."
- ⦿ **Kunden:** "Vent, kanskje jeg har det travelt og ikke trenger vinduet. Kanskje jeg burde lage en balkong".

I slike scenarier, som skjer ofte, fører selv en liten endring til en kjedereaksjon mellom flere systemer og roller. I dette tilfellet foregår nesten all kommunikasjon i den innledende fasen i tekstform: e-poster, chatter, møtereferater (fig. 4.4-1).

I et slikt tekstbasert kommunikasjonssystem for et byggeprosjekt er det svært viktig med et system for juridisk bekreftelse og registrering av alle datautvekslingsoperasjoner og alle beslutninger som tas. Dette for å sikre at alle beslutninger, instrukser og endringer er juridisk gyldige og sporbare, noe som reduserer risikoen for fremtidige "misforståelser"

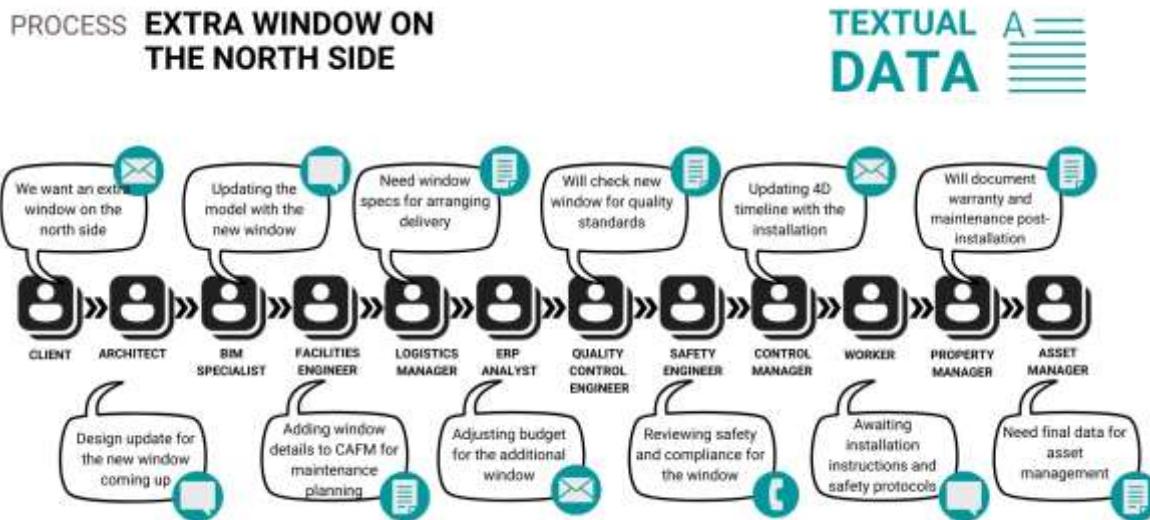


Fig. 4.4-1 Kommunikasjonen mellom byggherre og entreprenør i de innledende fasene av et prosjekt inneholder ofte tekstdata i flere formater.

Manglende juridisk kontroll og validering av beslutninger i de relevante systemene i et byggeprosjekt kan føre til alvorlige problemer for alle involverte. Enhver beslutning, ordre eller endring som tas uten skikkelig dokumentasjon og validering, kan føre til tvister (og rettssaker).

Juridisk konsolidering av alle beslutninger i tekstlig kommunikasjon kan bare sikres ved et stort antall signerte dokumenter, noe som vil falle på skuldrene til ledelsen, som er forpliktet til å registrere alle transaksjoner. Hvis hver deltaker må signere dokumenter for hver handling, mister systemet fleksibilitet og blir en byråkratisk labyrinth. Manglende transaksjonsbekreftelser vil ikke bare forsinke prosjektgjennomføringen, men kan også føre til økonomiske tap og forverring av forholdet mellom deltakerne, til og med juridiske problemer.

En slik transaksjonsgodkjenningsprosess, som vanligvis starter med tekstbaserte diskusjoner, utvikler seg gradvis til en dokumentutveksling i flere formater i de følgende trinnene (fig. 4.4-2), noe som kompliserer kommunikasjonen som tidligere bare foregikk via tekst, betraktelig. Uten klart definerte krav blir det nesten umulig å automatisere slike prosesser, som er fylt med data i flere formater og et stort antall tekniske krav.

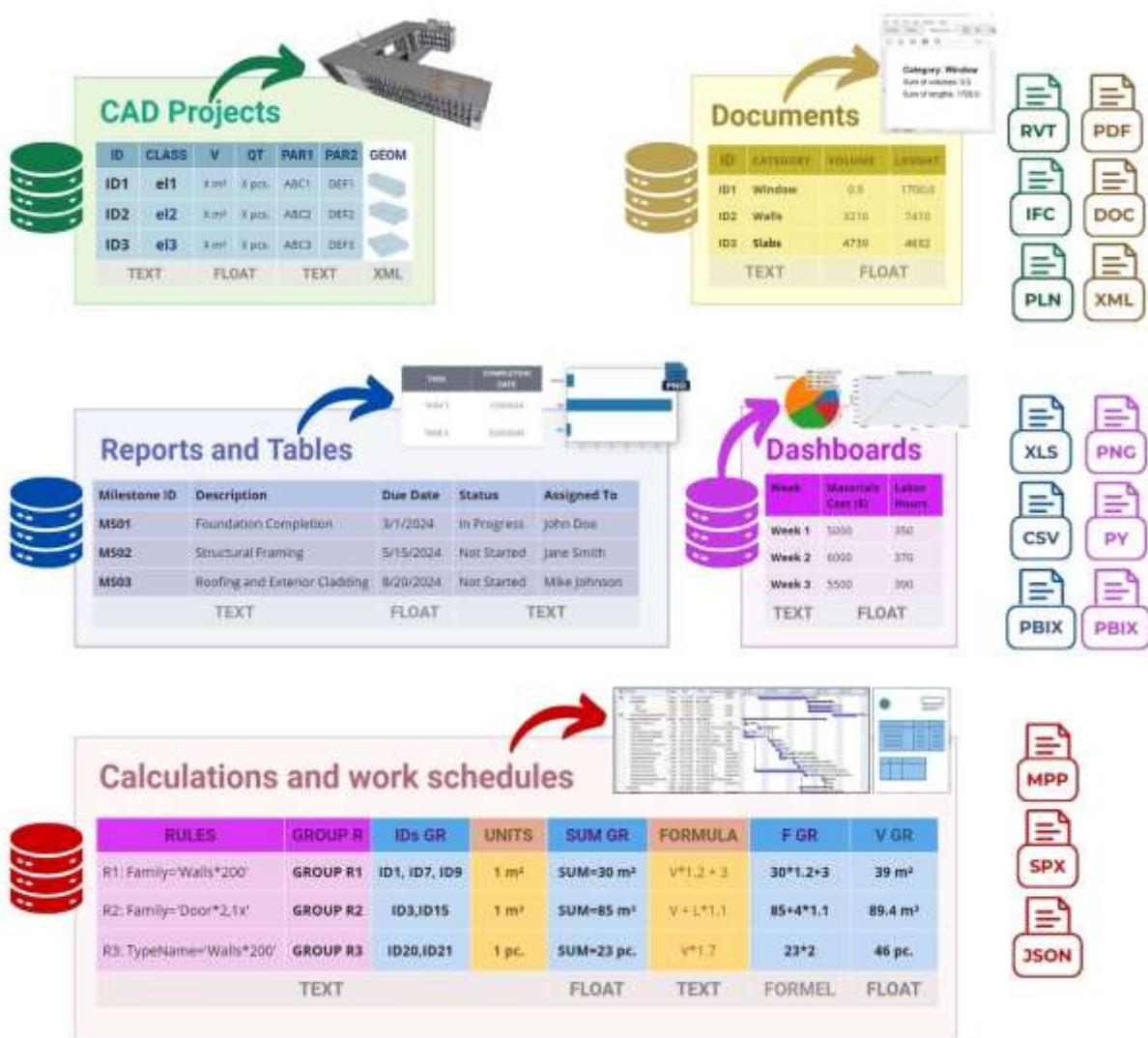


Fig. 4.4-2 Hvert system i byggefirmaets landskap fungerer som en kilde til juridisk relevante dokumenter i en rekke ulike formater.

Tekstkommunikasjon krever at hver enkelt fagperson enten setter seg inn i hele korrespondansen eller deltar regelmessig på alle møtene for å forstå den aktuelle statusen i prosjektet.

For å overvinne denne begrensningen er det nødvendig med en overgang fra tekstlig kommunikasjon til en strukturert kravmodell. Dette er bare mulig gjennom systematisk analyse, prosessvisualisering og beskrivelse av interaksjoner i form av flytdiagrammer og datamodeller (fig. 4.4-3). På samme måte som i datamodelleringen (fig. 4.3-7) beveget vi oss fra det kontekstuelle idénivået til det konseptuelle nivået ved å legge til de systemene og verktøyene som brukes av deltakerne, og koblingene mellom dem.

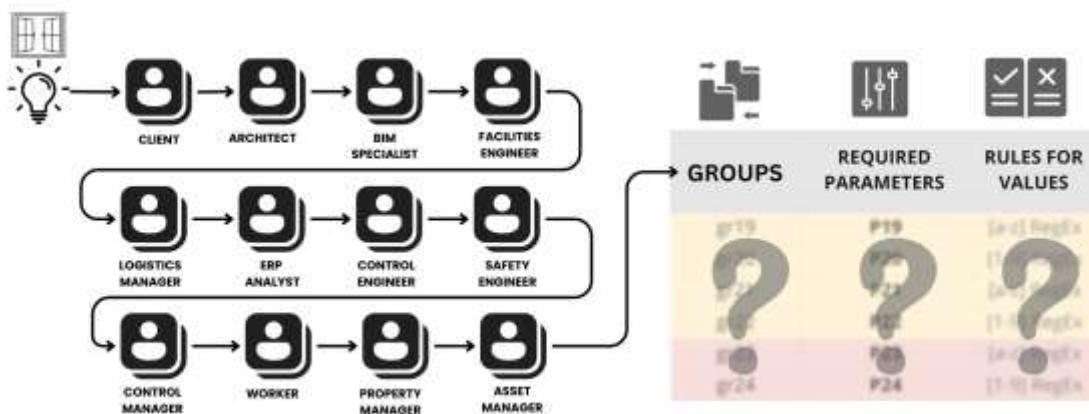


Fig. 4.4-3 For å lære hvordan man styrer og automatiserer valideringsprosessen, er det nødvendig å visualisere prosessene og strukturere kravene.

Det første steget i systematiseringen av krav og relasjoner er å visualisere alle koblinger og relasjoner ved hjelp av konseptuelle flytdiagrammer. Det konseptuelle nivået vil ikke bare gjøre det enklere for alle prosessdeltakerne å forstå hele prosesskjeden, men vil også tydelig vise hvorfor og for hvem dataene (og kravene) er nødvendige på hvert prosesstrinn.

Prosessflytskjemaer og effektiviteten til konseptuelle rammeverk

For å bygge bro over gapet mellom tradisjonelle og moderne tilnærminger til datahåndtering må bedriftene foreta et bevisst skifte fra fragmenterte tekstlige beskrivelser til strukturerte prosessrepresentasjoner. Utviklingen av data - fra leirtavler til digitale økosystemer - krever nye tenkeverktøy. Et slikt verktøy er konseptuell modellering ved hjelp av flytdiagrammer. Ved å lage visuelle diagrammer - flytdiagrammer, prosessdiagrammer, interaksjonsdiagrammer - kan prosjektdeltakerne innse hvordan deres handlinger og beslutninger påvirker hele beslutningssystemet.

Hvis prosessene ikke bare krever lagring av data, men også analyse eller automatisering av dem, må du begynne å skape et konseptuelt og visuelt kravlag.

I vårt eksempel (fig. 4.4-1) kan hver spesialist være en del av et lite team, men også av en større avdeling med opptil et dusin eksperter under kontroll av en daglig leder. Hver avdeling bruker en spesialisert applikasjonsdatabase (fig. 1.2-4, f.eks. ERP, CAD, MEP, CDE, ECM, CPM etc.), som regelmessig oppdateres med innkommende informasjon som er nødvendig for å opprette dokumenter, registrere den juridiske statusen til beslutninger og styre prosesser.

Transaksjonsprosessen ligner på arbeidet til antikkens ledere for 4000 år siden, da leirtavler og papyrus ble brukt til å bekrefte beslutninger juridisk. Forskjellen mellom moderne systemer og deres forgjengere av leire og papir er at moderne metoder i tillegg omfatter prosessen med å konvertere tekstlig informasjon til digital form for videre automatisk behandling i andre systemer og verktøy.

Ved å visualisere prosessen i form av konseptuelle flytdiagrammer kan man beskrive hvert trinn og samspillet mellom de ulike rollene, slik at en kompleks arbeidsflyt blir oversiktlig og enkel.

Visualisering av -prosesser sikrer at prosesslogikken er transparent og tilgjengelig for alle teammedlemmer.

Den samme kommunikative prosessen for å legge til et vindu i et prosjekt som ble beskrevet i form av tekst, meldinger (figur 4.4-1) og blokkdiagram, ligner på den konseptuelle modellen vi diskuterte i kapittelet om datamodellering (figur 4.4-4).

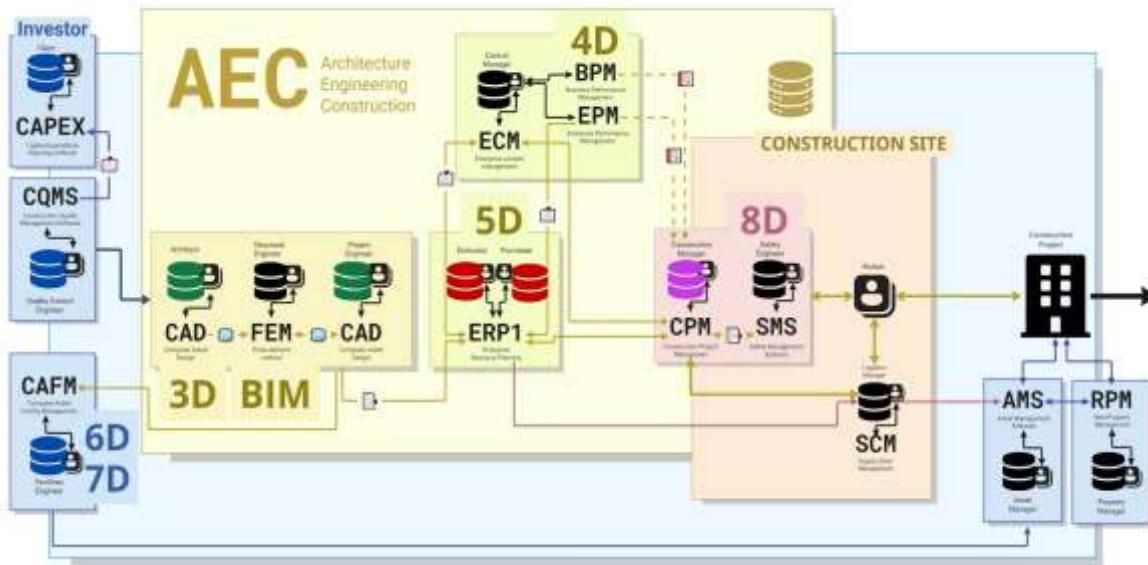


Fig. 4.4-4 Konseptdiagrammet viser prosjektdeltakerne som brukere av en database der forespørslene deres knytter sammen ulike systemer.

Selv om konseptuelle diagrammer er et viktig skritt, begrenser mange bedrifter seg til dette nivået og tror at et visuelt diagram er tilstrekkelig for å forstå prosessene. Dette skaper en illusjon av håndterbarhet: Ledere kan letttere oppfatte helheten i et slikt flytskjema og se sammenhengene mellom deltakere og stadier. Slike skjemaer gir imidlertid ikke et klart bilde av hvilke data som kreves for hver deltaker, i hvilket format de skal overføres, og hvilke parametere og attributter som er obligatoriske for automatisering. Et konseptuelt flytskjema er mer som et veikart: Det viser hvem som samhandler med hvem, men avslører ikke hva som overføres i disse interaksjonene.

Selv om en prosess er beskrevet i detalj på konseptuelt nivå ved hjelp av flytskjemaer, er det ingen garanti for at den er effektiv. Visualisering forenkler ofte arbeidet for ledere, slik at de letttere kan følge prosessen ved hjelp av trinnvis rapportering. For databaseingeniører er det imidlertid ikke sikkert at den konseptuelle fremstillingen er tydelig nok og gir en klar forståelse av hvordan prosessen skal implementeres på parameter- og kravnivå.

Etter hvert som vi beveger oss mot mer komplekse dataøkosystemer, blir den første implementeringen av konseptuelle og visuelle verktøy avgjørende for å sikre at dataprosessene ikke bare er effektive, men også

i tråd med organisasjonens strategiske mål. For å kunne oversette denne vindustilleggsprosessen (figur 4.4-1) til datakravnivået fullt ut, må vi gå et nivå dypere og oversette den konseptuelle visualiseringen av prosessen til det logiske og fysiske datanivået, de nødvendige attributtene og grenseverdiene for disse.

Strukturerte krav og regulære RegEx-uttrykk

Opp til 80% av dataene som skapes i bedrifter, er i ustukturerte eller halvstrukturerte formater [52] tekst, dokumenter, brev, PDF-filer, samtaler. Slike data (fig. 4.4-1) er vanskelige å analysere, verifisere, overføre mellom systemer og bruke i automatisering.

For å sikre håndterbarhet, gjennomsiktighet og automatisk validering er det nødvendig å oversette tekstlige og halvstrukturerte krav til veldefinerte, strukturerte formater. Strukturingsprosessen gjelder ikke bare dataene (som vi har diskutert i detalj i de første kapitlene i denne delen av boken), men også selve kravene, som prosjektdeltakerne vanligvis formulerer i fritekstform gjennom hele prosjektets livssyklus, ofte uten å tenke på at disse prosessene kan automatiseres.

På samme måte som vi allerede har konvertert data fra en ustukturert tekstlig form til en strukturert form, vil vi i kravarbeidsflyten konverte tekstlige krav til et strukturert "logisk og fysisk lag"-format.

Som i eksemplet med å legge til et vindu (figur 4.4-1), er neste trinn å beskrive datakravene i tabellform. Vi strukturerer informasjonen for hvert system som brukes av prosjektdeltakerne, ved å spesifisere nøkkelattributter og grenseverdier for disse

Ta for eksempel et slikt system (fig. 4.4-5) - Construction Quality Management System (CQMS), som brukes av kvalitetskontrollingeniøren på byggherresiden. Ved hjelp av systemet kontrollerer han om et nytt element i prosjektet - i dette tilfellet "nytt vindu" - er i samsvar med de etablerte standardene og kravene.

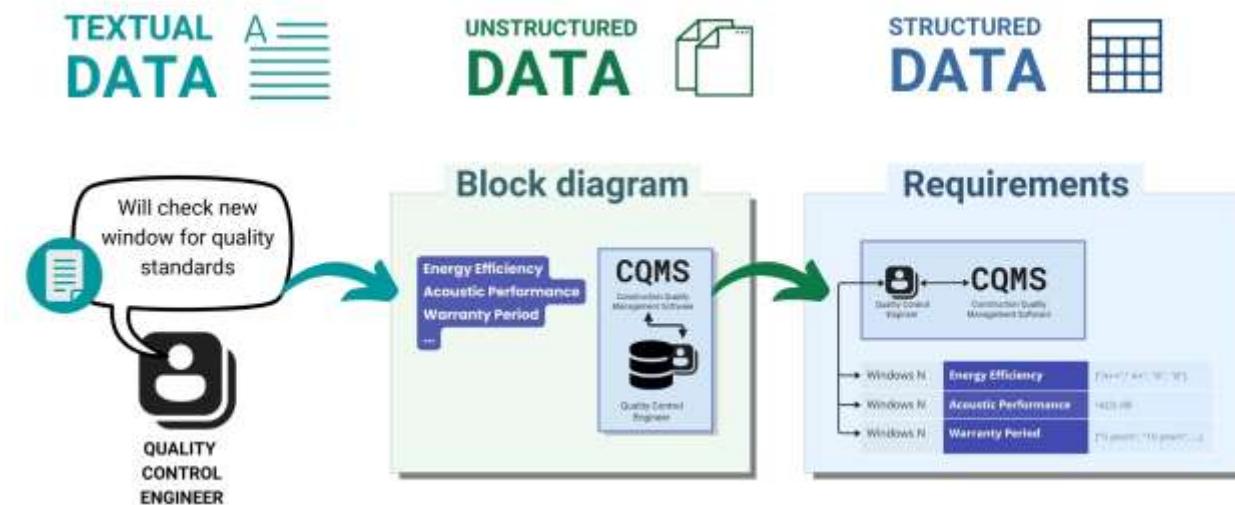


Fig. 4.4-5 Konvertering av tekstlige krav til et tabellformat med beskrivelser av entitetsattributter forenkler forståelsen for andre spesialister.

Ta for eksempel noen viktige krav for attributter for enheter av typen "vindussystemer" i CQMS -system (fig. 4.4-6): energieffektivitet, akustisk ytelse og garantiperiode. Hver kategori omfatter visse standarder og spesifikasjoner som må tas i betrakning ved utforming og installasjon av vindussystemer.



Fig. 4.4-6 Kvalitetskontrollingeniøren skal inspirere nye vinduselementer med hensyn til energieffektivitet, lydisolering og garantistandarder.

Datakravene som en kvalitetskontrollingeniør spesifiserer i form av en tabell, har for eksempel følgende grenseverdier:

- **Energieffektivitetsklassen for vinduer** går fra "A++", som angir den høyeste effektiviteten, til "B", som anses som det laveste akseptable nivået, og disse klassene er representert ved en liste over akseptable verdier **[A++, A+, A, A, A, A, B]**.
- **Vinduers akustiske isolasjonsevne**, målt i desibel og som viser deres evne til å redusere gatestøy, er definert ved **det** vanlige uttrykket $\Delta f(2)$ dB.

- **Attributtet "Warranty Period"** for entiteten "Window Type" starter på fem år, og angir denne perioden som minste tillatte periode ved valg av et produkt. Garantiperiodeverdier som **"5 år", "10 år" osv.]** eller den logiske betingelsen "**>5 (år)**" kan også angis.

I henhold til de innsamlede kravene, innenfor de etablerte attributtene, vil nye vinduskategorier eller klasseelementer med karakterer under "B", for eksempel "C" eller "D", ikke bestå energieffektivitetstesten. Akustisk vindusisolasjon i data eller dokumenter til kvalitetssikringsingenøren skal merkes med et tosifret tall etterfulgt av postfiksset "dB", for eksempel "35 dB" eller "40 dB", og verdier utenfor dette formatet, for eksempel "9 D B" eller "100 desibel", vil ikke bli akseptert (da de ikke vil passere mønsteret for RegEx-strenger). Garantiperioden må begynne med minimum "5 år", og vinduer med kortere garantiperioder, for eksempel "3 år" eller "4 år", vil ikke oppfylle kravene som kvalitetssingenøren har beskrevet i tabellformatet.

For å sjekke slike attributparameterverdier mot grenseverdier fra krav i valideringsprosessen bruker vi enten en liste over tillatte verdier ([A, B, C]), ordbøker ([A: H1, H2; B: W1, W2]), logiske operasjoner (f.eks. ">", "<", "<=", ">=", "==") for numeriske verdier) og regulære uttrykk (for streng- og tekstverdier, som i attributtet "Acoustic Performance"). Regulære uttrykk er et svært viktig verktøy når du arbeider med strengverdier.

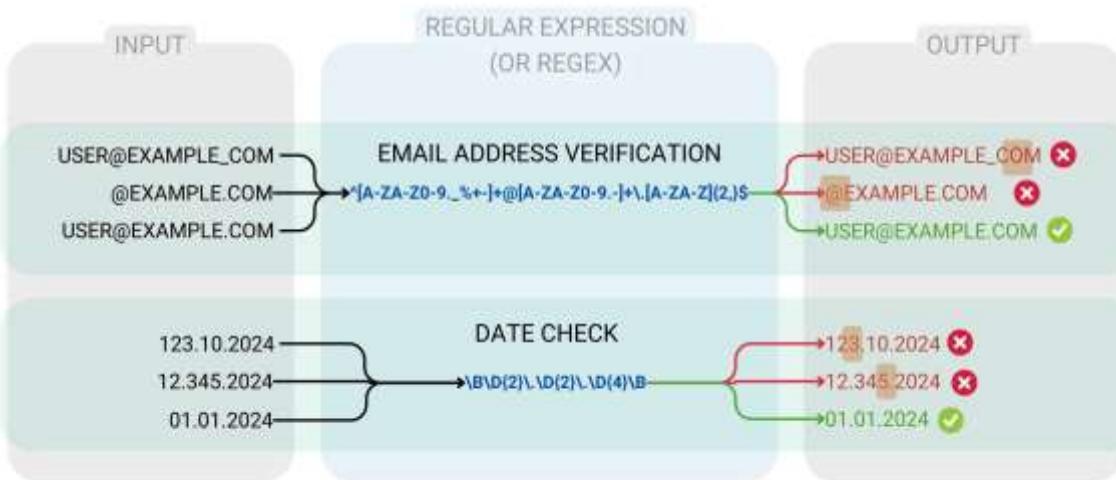
Regulære uttrykk (RegEx) brukes i programmeringsspråk, inkludert Python (Re-biblioteket), for å finne og modifisere strenger. Regex er som en detektiv i strengverdenen, og kan identifisere tekstmønstre i tekst med stor presisjon.

I regulære uttrykk beskrives bokstaver direkte ved hjelp av de tilsvarende tegnene i alfabetet, mens tall kan representeres ved hjelp av spesialtegnet \d, som tilsvarer et hvilket som helst siffer fra 0 til 9. Firkantede parenteser brukes for å angi et bokstav- eller sifferområde, f.eks. [a-z] for alle små bokstaver i det latinske alfabetet eller [0-9], som tilsvarer \d. For ikke-numeriske tegn og tegn som ikke er bokstaver, brukes henholdsvis \D og \W.

Populære RegEx-brukstilfeller (figur 4.4-7):

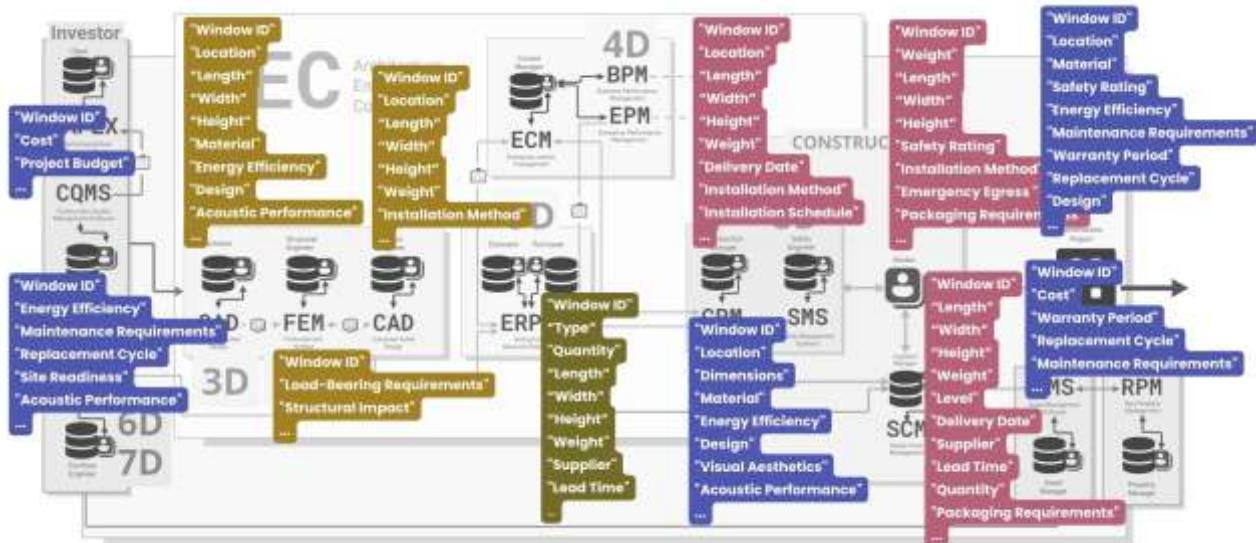
- **Verifisering av e-postadresse:** For å sjekke om en streng er en gyldig e-postadresse, kan du bruke malen "`^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$`".
- **Datoekstraksjon:** "`\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b`"-malen kan brukes til å trekke ut dato fra tekst i DD.MM.YYYY-format.
- **Verifisering av telefonnumre:** For å verifisere telefonnumre i formatet +49(000)000-0000, vil mønsteret se ut som "`^+\d{2}(\d{3})\d{3}-\d{4}`".

Ved å oversette kravene fra en kvalitetssikringsingenør til formatet til attributter og deres grenseverdier (fig. 4.4-6), har vi forvandlet dem fra deres opprinnelige tekstformat (samtaler, brev og forskriftsdokumenter) til en organisert og strukturert tabell, noe som gjør det mulig å automatisk kontrollere og analysere alle innkommende data (f.eks. nye -elementer i kategorien Vindu). Ved hjelp av krav kan data som ikke er kontrollert, automatisk forkastes, og de kontrollerte dataene kan automatisk overføres til systemene for videre behandling.



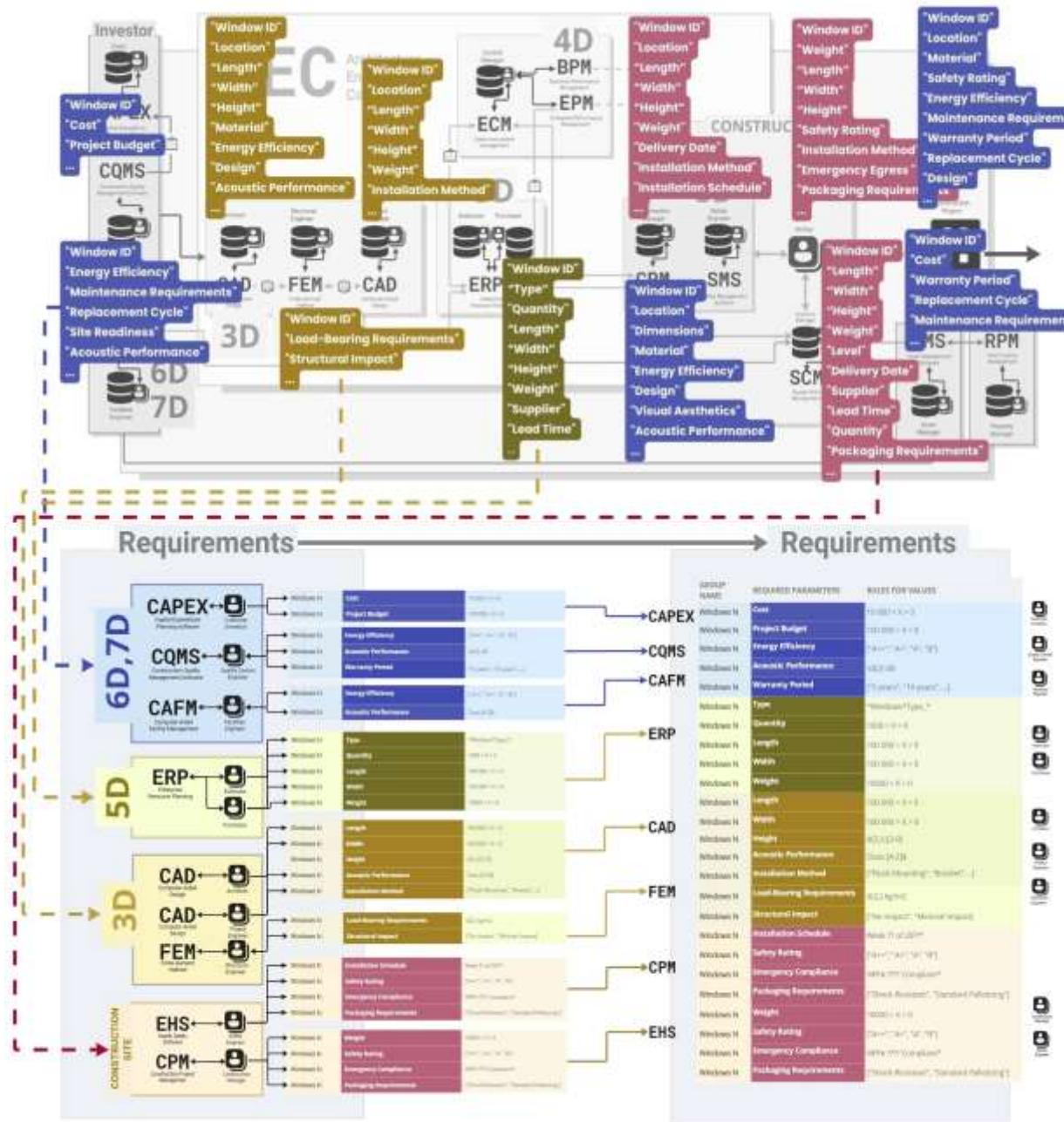
Figur 4.4-7 Bruk av regulære uttrykk er et svært viktig verktøy i valideringsprosessen for tekstdata

Når vi nå går fra det konseptuelle til det logiske nivået i arbeidet med krav, konverterer vi alle kravene til alle spesialistene i prosessen med å installere et nytt vindu (fig. 4.4-4) til en ordnet liste i attributtformat og legger til disse listene med de nødvendige attributtene i flytskjemaet vårt for hver spesalist (fig. 4.4-8).



Figur 4.4-8 På det logiske prosessnivået legges attributtene som håndteres av hver spesalist, inn i deres respektive systemer.

Ved å legge alle attributtene inn i en felles prosesstabell forvandler vi informasjonen som tidligere ble presentert som tekst og dialog på det konseptuelle nivået (fig. 4.4-1), til den strukturerte og systematiske formen til tabeller på det fysiske nivået (fig. 4.4-9).



Figur 4.4-9 Konvertering av ustrukturert fagdialog til strukturerte tabeller gjør det lettere å forstå krav på det fysiske nivået.

Nå må datakravene kommuniseres til spesialistene som lager informasjon for spesifikke systemer. Hvis du for eksempel jobber i en CAD-database, bør du samle inn alle nødvendige parametere basert på sluttbruksscenarioene for dataene før du begynner å modellere elementer. Dette starter vanligvis med driftsfasen, etterfulgt av byggeplassen, logistikkavdelingen, kalkulasjonsavdelingen, avdelingen for strukturberegninger og så videre. Først etter at du har tatt hensyn til kravene fra alle disse områdene, kan du begynne å skape data - basert på de innsamlede parameterne. På denne måten kan du i fremtiden automatisere verifiseringen og overføringen av data gjennom hele kjeden.

Når nye data oppfyller kravene, integreres de automatisk i bedriftens dataøkosystem og går direkte til de brukerne og systemene de er beregnet på. Verifisering av data mot attributter og deres verdier sikrer at informasjonen oppfyller de nødvendige kvalitetsstandardene og er klar til bruk i bedriftens scenarier.

Datakravene er definert, og før verifiseringen kan begynne, må dataene som skal verifiseres, opprettes, innhentes eller samles inn, eller den aktuelle informasjonstilstanden i databasene må registreres for å kunne brukes i verifiseringsprosessen.

Datainnsamling for verifiseringsprosessen

Før valideringen kan begynne, er det viktig å sørge for at dataene er tilgjengelige i en form som er egnet for valideringsprosessen. Dette betyr ikke bare å ha informasjonen tilgjengelig, men å forberede den: Dataene må samles inn og transformeres fra ustrukturerte, løst strukturerte, tekstlige og geometriske formater til en strukturert form. Denne prosessen er beskrevet i detalj i de foregående kapitlene, der metoder for å transformere ulike typer data ble diskutert. Som et resultat av alle transformasjonene tar de innkommende dataene form av åpne, strukturerte tabeller (fig. 4.1-2, fig. 4.1-9, fig. 4.1-13).

Med kravene og strukturerte tabeller med de nødvendige parameterne og grenseverdiene (fig. 4.4-9) kan vi begynne å validere dataene - enten som én automatisert prosess (Pipeline) eller som en trinnvis validering av hvert enkelt innkommende dokument.

For å kunne starte kontrollen må man enten motta en ny fil som input eller fastsette dataenes nåværende tilstand - lage et øyeblikksbilde eller eksportere nåværende og innkommende data, eller sette opp en forbindelse til en ekstern eller intern database. I det aktuelle eksemplet opprettes et slikt øyeblikksbilde ved automatisk å konvertere CAD-data fra til et strukturert format, for eksempel kl. 23:00:00 fredag 29. mars 2024, etter at alle konstruktørene har gått hjem.



Figur 4.4-10 CAD-databasens øyeblikksbilde (BIM) som viser gjeldende attributtinformasjon for en ny entitet av klassen "Window" i den gjeldende versjonen av prosjektmodellen.

Takket være reverse engineering-verktøyene som er omtalt i kapittelet "Oversette DAK-data (BIM) til en strukturert form", kan denne informasjonen fra ulike DAK-verktøy og BIM-editorer organiseres i separate tabeller (fig. 4.4-11) eller kombineres i en felles tabell som kombinerer ulike deler av prosjektet (fig. 9.1-10).

En slik tabell - database viser unike identifikatorer for vinduer og dører (ID-attributt), typenavn (TypeName), dimensjoner (Width, Length), materialer (Material), samt indikatorer for energi- og akustisk ytelse og andre

egenskaper. En slik tabell fylt ut i CAD-programmet (BIM) samles inn av en designingeniør fra ulike avdelinger og dokumenter, og danner en informasjonsmodell av prosjektet.

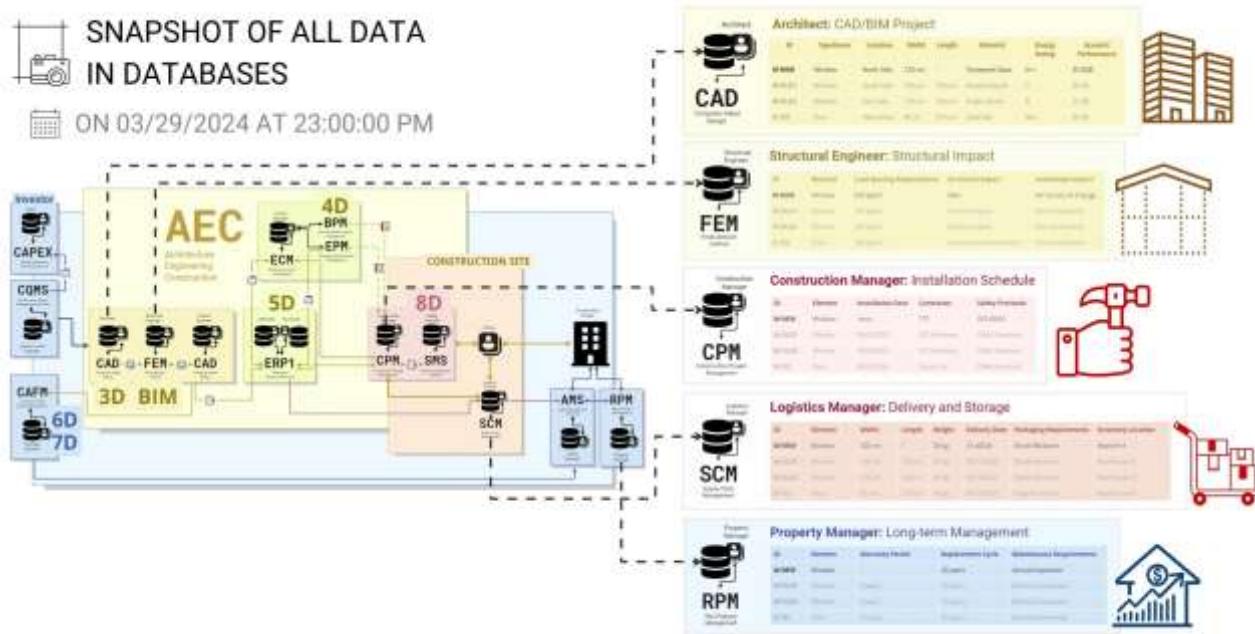


Figur 4.4-11 Strukturerte data fra CAD systemer kan være en todimensjonal tabell med kolonner som angir attributter for elementer.

Ekte CAD (BIM) prosjekter omfatter titusener eller hundretusener av elementer (fig. 9.1-10). Elementer i CAD-formater kategoriseres automatisk etter type og kategori - fra vinduer og dører til plater, dekker og veggger. Unike identifikatorer (f.eks. native ID, som settes automatisk av DAK-løsningen) eller typeattributter (typenavn, type, familie) gjør det mulig å spore det samme objektet i forskjellige systemer. For eksempel kan et nytt vindu på nordveggen i en bygning identifiseres entydig med en enkelt identifikator "W-NEW" i alle relevante systemer i organisasjonen.

Selv om enhetsnavn og identifikatorer bør være konsistente på tvers av alle systemer, kan settet av attributter og verdier som er knyttet til disse enhetene, variere betydelig avhengig av brukskonteksten. Arkitekter, bygningsingeniører, fagfolk innen konstruksjon, logistikk og eiendomsvedlikehold oppfatter alle de samme elementene på forskjellige måter. Hver av dem bruker sine egne klassifiseringsmetoder, standarder og mål: Noen ser på et vindu ut fra et rent estetisk synspunkt og vurderer form og proporsjoner, mens andre ser på det ut fra et teknisk eller driftsmessig synspunkt og analyserer varmeledningsevne, installasjonsmetode, vekt eller vedlikeholds krav. Når man modellerer data og beskriver elementer, er det derfor viktig å ta hensyn til hvor allsidig de kan brukes, og å sikre at dataene er konsistente, samtidig som man tar hensyn til bransjespesifikke forhold.

For hver rolle i bedriftens prosesser finnes det spesialiserte databaser med eget brukergrensesnitt - fra design og beregninger til logistikk, installasjon og bygningsdrift (fig. 4.4-12). Hvert slikt system administreres av et profesjonelt team av spesialister gjennom et spesielt brukergrensesnitt eller gjennom databaseforespørslar, og bak summen av alle beslutninger som tas på grunnlag av de innlagte verdiene i enden av kjeden, står systemlederen eller avdelingslederen, som er ansvarlig for at de innlagte dataene er juridisk gyldige og av god kvalitet overfor sine motparter som betjener andre systemer.



Figur 4.4-12 Samme entitet har samme identifikator i ulike systemer, men ulike attributter som bare er viktige i det aktuelle systemet.

Når vi har organisert innsamlingen av strukturerte krav og data på logisk og fysisk nivå, gjenstår det å sette opp en prosess som automatisk kontrollerer dataene fra ulike innkommende dokumenter og ulike systemer opp mot de tidligere innsamlede kravene.

Verifikasiing av data og resultater av verifikasiingen

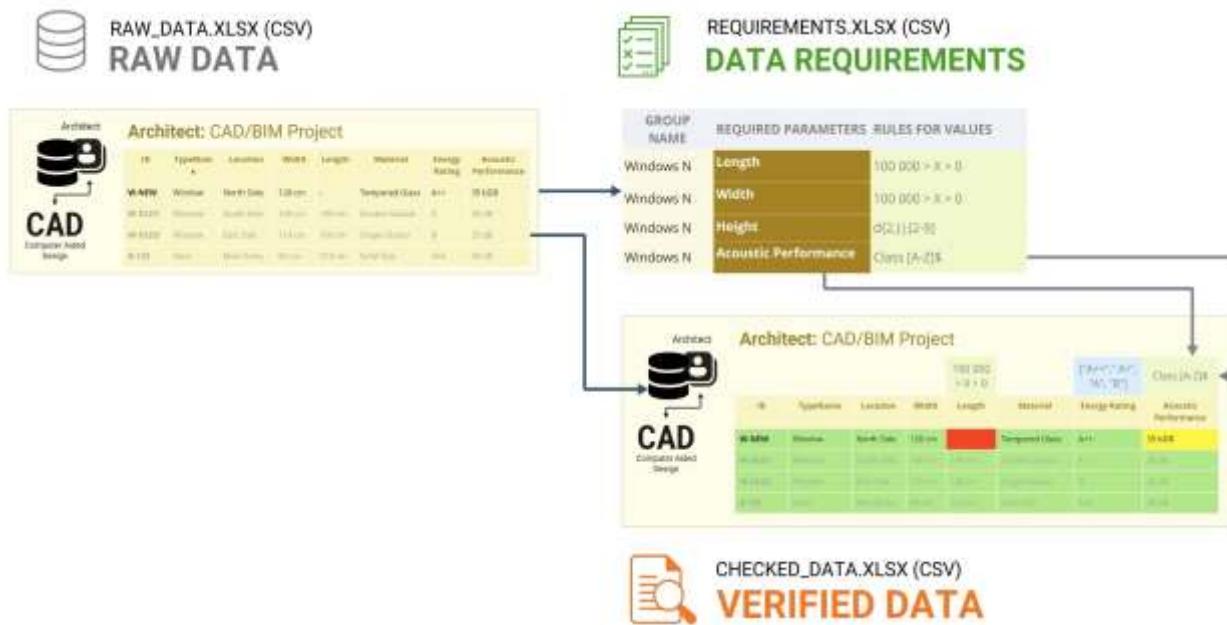
Alle nye data som kommer inn i systemet - det være seg dokumenter, tabeller eller databaseoppføringer fra kunden, arkitekten, ingeniøren, arbeidslederen, logistikeren eller eiendomsforvalteren - må valideres mot de kravene som er formulert tidligere (fig. 4.4-9). Valideringsprosessen er kritisk: Eventuelle feil i dataene kan føre til feilberegninger, forsinkelser i tidsplanen og til og med økonomiske tap. For å minimere slike risikoer bør man organisere en systematisk og repeterbar, iterativ datavalidiseringsprosedyre.

For å validere nye data som kommer inn i systemet - ustukturerte, tekstlige eller geometriske - må de konverteres til et løst strukturert eller strukturtformat. Valideringsprosessen må deretter kontrollere dataene mot en komplett liste over påkrevde attributter og deres tillatte verdier.

Konvertering av ulike typer data: tekst, bilder, PDF-dokumenter og blandede CAD (BIM) data til en strukturert form ble diskutert i detalj i kapittelet "Konvertering av data til en strukturert form".

Et eksempel er en tabell hentet fra et CAD (BIM)-prosjekt (fig. 4.4-11). Den inneholder halvstrukturerte geometriske data og strukturert attributtinformasjon om prosjektets enheter (fig. 3.1-14) - f.eks. et element fra klassen "Windows".

For å utføre valideringen sammenligner vi attributtverdiene (fig. 4.4-11) med referansegrenseverdiene som er definert av eksperter i form av et krav (fig. 4.4-9). Den endelige sammenligningstabellen (fig. 4.4-13) viser hvilke verdier som er akseptable, og hvilke som må korrigeres før dataene kan brukes utenfor CAD-applikasjoner (BIM).



Figur 4.4-13 Den endelige valideringstabellen fremhever de attributtverdiene for den nye entiteten av klassen "Windows" som du bør være oppmerksom på.

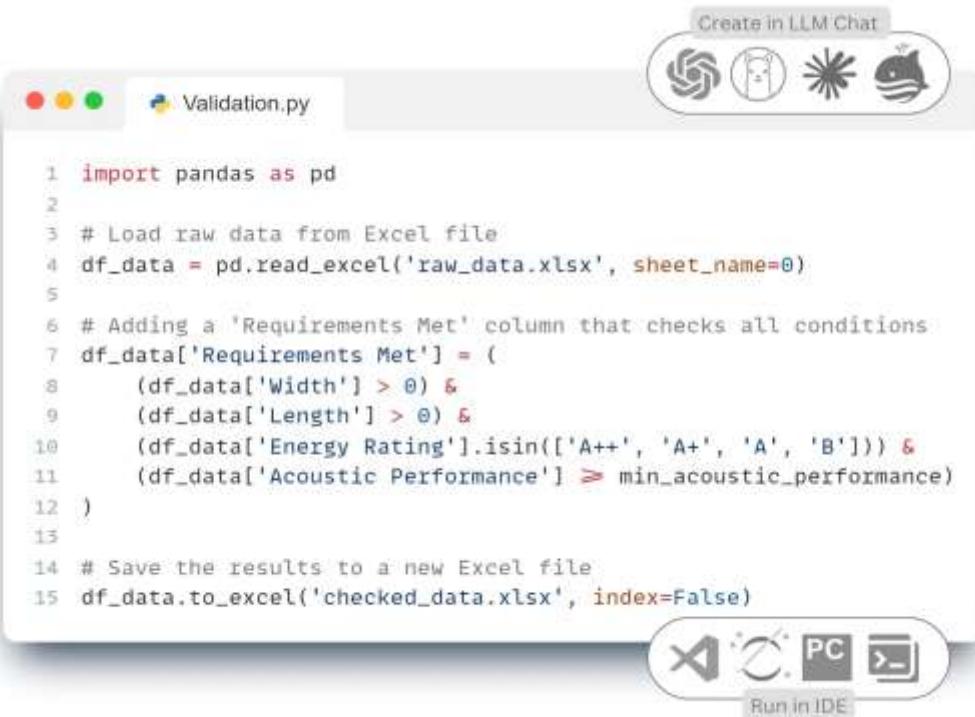
Ved å implementere en lignende løsning ved hjelp av Pandas-biblioteket, som vi beskrev tidligere i kapittelet "Pandas: Et uunnværlig verktøy for dataanalyse", vil vi validere data fra en tabellarisk fil hentet fra en CAD-fil (BIM) (RVT, IFC, DWG, NWS, DGN) (figur 4.4-11) ved hjelp av kravene fra en annen tabellarisk kravfil (figur 4.4-9).

For å få koden må vi beskrive i ledeteksten for LLM at vi må laste inn dataene fra filen **raw_data.xlsx** (et komplett sett med data fra CAD-databasen (BIM)), **kontrollere** dem og lagre resultatet i en ny fil **checked_data.xlsx** (fig. 4.4-13).

- La oss få koden ved hjelp av LLM uten å nevne Pandas-biblioteket:

Skriv kode for å sjekke tabellen fra filen **raw_data.xlsx** og validere dem ved hjelp av følgende valideringsregler: verdiene i kolonnene "Width" og "Length" er større enn null, "Energy Rating" er inkludert i listen ["A++", "A+", "A", "B"], og "Acoustic Performance" er en variabel som vi vil spesifisere senere - med den siste valideringskolonnen lagt til, og lagre den endelige tabellen i en ny Excel-fil **checked_data.xlsx**. ↵

- LLMs svar vil beskrive et kort eksempel på Python-kode som kan foredles og utvides ved hjelp av påfølgende spørsmål :



```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

Figur 4.4-14 Koden som genereres av LLM- -modellen, kontrollerer det konverterte CAD-designet (BIM) mot attributtkravene i form av grenseverdier.

Koden som genereres av LLM-språkmodellen, kan brukes i alle populære IDE-er eller nettbaserte verktøy: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære nettbaserte verktøy som Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Utførelse av koden (fig. 4.4-14) vil vise at "entitetselementene" W-OLD1, W-OLD2, D-122 (og andre -elementer) fra CAD-databasen (BIM) oppfyller attributtkravene: bredde og lengde er større enn null, og energieffektivitetsklassen er en av listeverdiene "A++", "A", "B", "C" (fig. 4.4-15).

Elementet W-NEW, som vi trenger og nylig har lagt til, og som er ansvarlig for den nye elementklassen "Window" på nordsiden, er ikke i samsvar med kravene (attributtet "Requrmets Met") fordi lengden er null (en verdi på "0,0" anses som uakseptabel i henhold til "Width">>0-regelen vår), og fordi det ikke angir en energieffektivitetsklasse.



CHECKED_DATA.XLSX (CSV)

VERIFIED DATA

ID	Type Name	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met	
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B-	30	True

Figur 4.4-15 Checking identifiserer enheter som ikke har gått gjennom verifiseringsprosessen, og legger til et nytt attributt med verdiene "False" eller "True" i resultatene.

På samme måte kontrollerer vi at alle prosjektelelementene (entitetene) og de nødvendige attributtene for hvert av systemene, tabellene eller databasene er konsistente i alle dataene vi mottar fra ulike spesialister (figur 4.4-1) i løpet av prosessen med å legge til et vindu i prosjektet.

I den siste tabellen er det praktisk å markere resultatene av kontrollen med farger for visualisering: attributter som har bestått kontrollen er markert med grønt, gult - verdier med ikke-kritiske avvik, og rødt - kritiske uoverensstemmelser (fig. 4.4-16).

Som et resultat av valideringen (figur 4.4-16) får vi en liste over pålitelige og validerte elementer med tilhørende identifikatorer som har blitt verifisert for å oppfylle attributtkravene. Validerte elementer gir sikkerhet for at disse elementene oppfyller de angitte standardene og spesifikasjonene for alle systemer som er involvert i prosessen med å legge til elementer i Window-klassen eller en hvilken som helst annen klasse (vi vil snakke mer om automatisering av datavalidering og oppretting av en automatisert ETL-prosess i kapittelet "Automatisering av ETL og datavalidering").

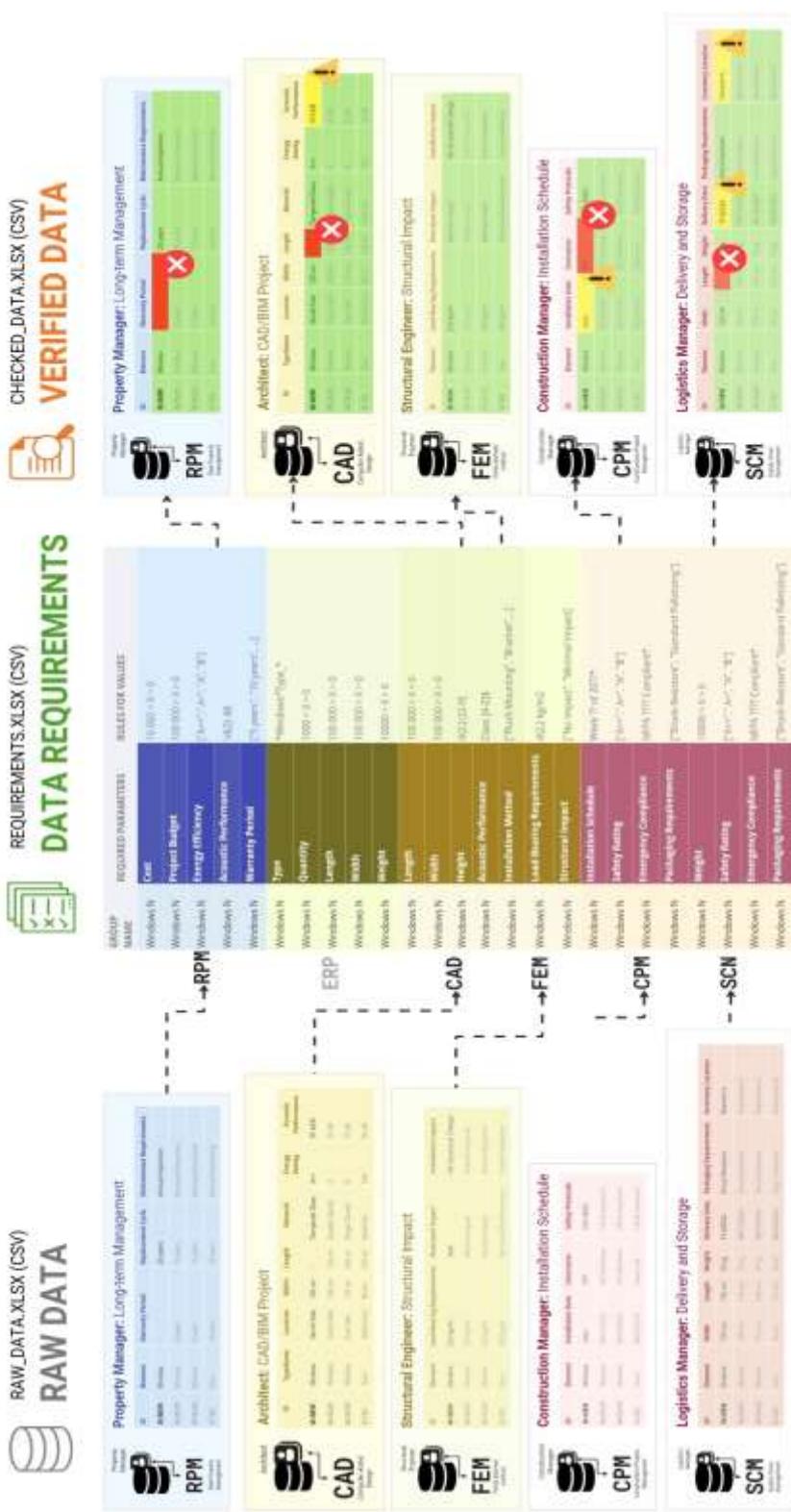


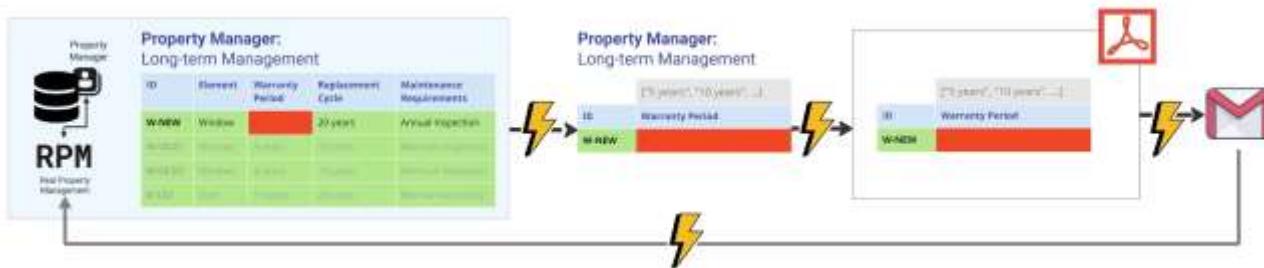
Fig. 4.4-16 Resultatet av kontrollen som er utført for alle systemer, gjør det mulig å fastslå hvilke data som ikke oppfyller bedriftens krav.

Enheter som har blitt validert, krever vanligvis ikke mye oppmerksomhet. De går videre til neste trinn i bearbeidingen og integreringen i andre systemer uten hindringer. I motsetning til "kvalitetselementer" er det de elementene som ikke blir validert, som er av størst interesse. Informasjon om slike avvik er avgjørende: Den bør ikke bare formidles i tabellrapporter, men også ved hjelp av ulike visualiseringsverktøy. En grafisk fremstilling av resultatene av verifiseringen bidrar til at man raskt kan vurdere den generelle tilstanden for datakvaliteten, identifisere problemområder og raskt iverksette korrigende eller avhjelpende tiltak.

Visualisering av verifiseringsresultater på

Visualisering er et viktig verktøy for å tolke inspeksjonsresultatene. I tillegg til de vanlige oppsummeringstabellene kan det omfatte dashboard, diagrammer og automatisk genererte PDF-dokumenter som grupperer prosjektelementer etter inspeksjonsstatus. Fargekoding kan spille en viktig rolle her: grønt kan indikere elementer som har blitt validert, gult kan indikere elementer som krever ekstra oppmerksomhet, og rødt kan indikere elementer som har kritiske feil eller mangler nøkkeldata.

I vårt eksempel (fig. 4.4-1) analyserer vi data fra hvert system trinn for trinn: fra CAD (BIM) og eiendomsforvaltning til logistikk og installasjonsplaner (fig. 4.4-16). Etter revisjonen genereres det automatisk individuelle varsler eller rapportdokumenter for hver spesialist, f.eks. i PDF-format (fig. 4.4-17). Hvis dataene er korrekte, mottar spesialisten en kort melding: "Takk for samarbeidet". Hvis det oppdages avvik, sendes det en detaljert rapport med følgende ordlyd: "Dette dokumentet viser elementer, deres identifikatorer, attributter og verdier som ikke har blitt kontrollert for samsvar".



Figur 4.4-17 Validering og automatisk rapportgenerering gjør det raskere å finne og forstå datamangler for den som skaper dataene.

Takket være den automatiserte valideringsprosessen - så snart en feil eller et datagap oppdages, sendes det umiddelbart et varsel i form av en chatmelding, e-post eller PDF-dokument til personen som er ansvarlig for å opprette eller behandle de aktuelle entitetene og attributtene deres (fig. 4.4-18), med en liste over elementer og attributtbeskrivelser som ikke er validert.



Fig. 4.4-18 Automatiske inspeksjonsrapporter gjør det lettere å forstå feil og gjør arbeidet med å komplettere prosjektdataene raskere.

Hvis for eksempel et dokument kommer inn i eiendomsforvaltingssystemet (etter strukturering) med attributtet "Garantiperiode" feil utfylt, mottar eiendomsforvalteren et varsel med en liste over attributter som må kontrolleres og korrigeres.

På samme måte vil eventuelle mangler i installasjonsplanen eller logistikkdataene føre til at det automatisk genereres en rapport, og at det for eksempel sendes et chat-varsle eller en e-post med resultatet av inspeksjonen til den aktuelle spesialisten.

I tillegg til PDF -dokumenter og grafer med resultater, er det mulig å lage dashbord og interaktive 3D - modeller (figur 7.1-6, figur 7.2-12) som fremhever elementer med manglende attributter, slik at brukerne visuelt kan bruke 3D-geometrier av elementer til å filtrere og evaluere kvaliteten og fullstendigheten av disse elementene i prosjektet.

Visualisering av inspeksjonsresultater på i form av automatisk genererte dokumenter, grafer eller dashbord forenkler tolkningen av data og legger til rette for effektiv kommunikasjon mellom prosjektdeltakerne.

Prosessens med automatisk kontroll av data fra ulike systemer og informasjonskilder kan sammenlignes med informert beslutningstaking i hverdagen. På samme måte som selskaper i byggebransjen tar hensyn til mange variabler - fra påliteligheten til inngangsdataene til deres innvirkning på tidspunkt, kostnader og kvalitet i prosjektgjennomføringen - veier man også en hel rekke faktorer mot hverandre når man skal ta viktige beslutninger, for eksempel når man skal velge et sted å bo: transporttilgjengelighet, infrastruktur, kostnader, sikkerhet og livskvalitet. Alle disse avveiningene danner et system av kriterier som ligger til grunn for de endelige beslutningene som utgjør våre liv.

Sammenligning av datakvalitetssjekker med menneskers livsbehov

Til tross for den stadige utviklingen av metoder og verktøy for datakvalitetskontroll, er det grunnleggende prinsippet om informasjonssamsvar uforandret. Dette prinsippet er bygget inn i fundamentet for et velutviklet styringssystem, enten det er i forretningslivet eller i hverdagen.

Prosessens med iterativ datavalidering ligner mye på beslutningsprosessen som alle står overfor hver dag. I begge tilfeller er vi avhengige av erfaring, data og ny informasjon etter hvert som den blir tilgjengelig. Og stadig flere beslutninger i livet og yrkeslivet - fra strategiske til hverdaglige avgjørelser - blir tatt på grunnlag av data.

Når vi for eksempel skal velge et sted å bo eller en livspartner, danner vi oss intuitivt en tabell med kriterier og egenskaper som vi sammenligner alternativene med (fig. 4.4-19). Disse egenskapene - det være seg personlige egenskaper hos en person eller parametere ved en eiendom - representerer attributter som påvirker den endelige beslutningen.

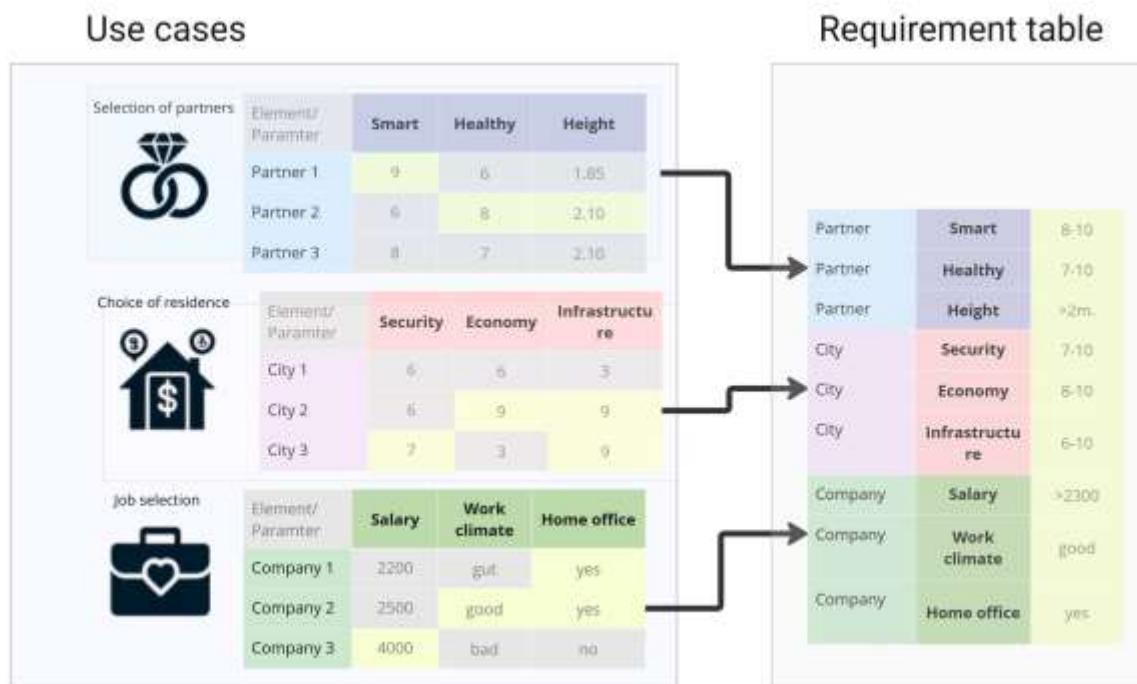


Fig. 4.4-19 Valget av bosted, arbeid eller partnerskap er basert på individuelle krav til attributter.

Bruk av strukturerte data og en formalisert tilnærming til å beskrive krav (figur 4.4-20) bidrar til mer informerte og velbegrunnede valg både i yrkeslivet og privatlivet.

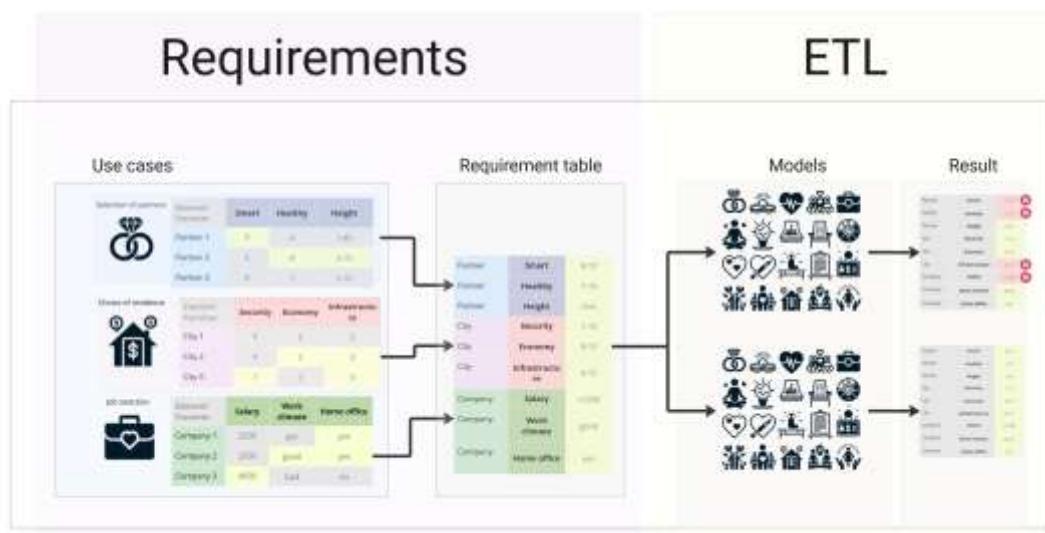


Fig. 4.4-20 Formalisering av krav gjør det mulig å systematisere livsoppfatninger og forretningsbeslutninger.

Datadrevet beslutningstaking er ikke utelukkende et forretningsverktøy. Den er sømløst integrert i hverdagen også, og følger vanlige databehandlingstrinn (fig. 4.4-21) som ligner på ETL-prosessen (Extract, Transform, Load) som vi allerede har diskutert i begynnelsen av denne delen når det gjelder strukturering av data, og som vi skal se nærmere på i forbindelse med oppgaveautomatisering i bokens syvende del:

- **Data som grunnlag (Utdrag):** På alle områder - enten det gjelder arbeid eller privatliv - samler vi inn informasjon. I forretningslivet kan det være rapporter, tall, markedsdata; i privatlivet kan det være personlige erfaringer, råd fra nære og kjære, tilbakemeldinger, observasjoner.
- **Evalueringskriterier (Transform):** Den innsamlede informasjonen tolkes på grunnlag av forhåndsdefinerte kriterier. På jobben er dette resultatindikatorer (KPI-er), budsjettbegrensninger og normer; i privatlivet er det parametere som pris, bekvemmelighet, pålitelighet, karisma osv.
- **Prediksjon og risikoanalyse (Load):** Den siste fasen innebærer beslutningstaking basert på analyse av de transformerte dataene og sammenligning av mulige konsekvenser. Dette ligner på forretningsprosesser, der data går gjennom et forretningslogikk- og risikofilter.

Beslutningene vi tar - fra trivielle preferanser som hva vi skal spise til frokost til store livshendelser som valg av karriere eller livspartner - er i seg selv et resultat av bearbeiding og evaluering av data.

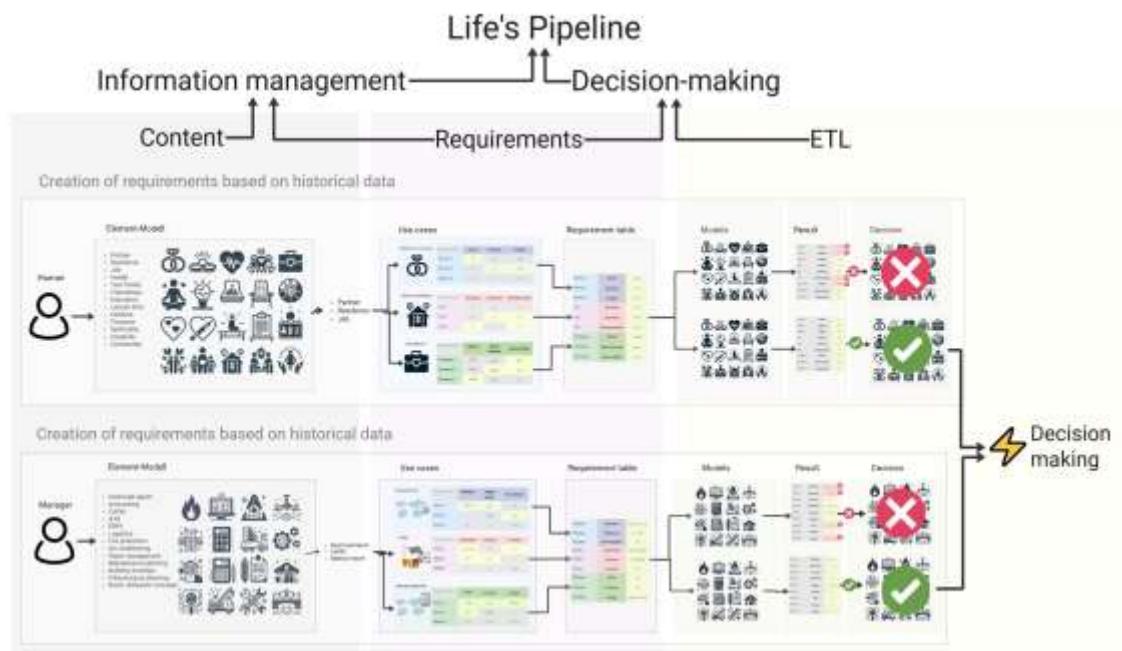


Fig. 4.4-21 Næringslivet og livet generelt består av en rekke datadrevne beslutninger der kvaliteten på dataene som brukes til å ta beslutninger, er en nøkkelfaktor.

Alt i livene våre henger sammen, og på samme måte som levende organismer, inkludert mennesker, følger naturlovene, utvikler seg og tilpasser seg skiftende forhold, gjenspeiler menneskelige prosesser, inkludert hvordan vi samler inn og analyserer data, disse naturlige prinsippene. Det tette forholdet mellom naturen og menneskelig aktivitet bekrefter ikke bare vår avhengighet av naturen, men også vårt ønske om å bruke lover som er utviklet gjennom millioner av år med evolusjon, til å skape dataarkitekturen, prosesser og systemer for beslutningstaking.

Ny teknologi, særlig innen bygg og anlegg, er et godt eksempel på hvordan menneskeheden gang på gang lar seg inspirere av naturen til å skape bedre, mer bærekraftige og effektive løsninger.

Neste trinn: gjøre data om til nøyaktige beregninger og planer

I denne delen har vi sett på hvordan man konverterer ustrukturerte data til et strukturert format, utvikler datamodeller og organiserer prosesser for å kontrollere kvaliteten på informasjonen i byggeprosjekter. Datahåndtering, standardisering og klassifisering er en grunnleggende prosess som krever en systematisk tilnærming og en klar forståelse av forretningskravene. Teknikkene og verktøyene som omtales i denne delen, muliggjør pålitelig integrering mellom ulike systemer gjennom hele livssyklusen til et objekt.

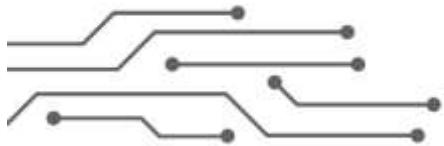
For å oppsummere denne delen, la oss trekke frem de viktigste praktiske trinnene som vil hjelpe deg med å bruke de diskuterte tilnærmingene i det daglige arbeidet:

- Begynn med å systematisere kravene
- Opprett et register med attributter og parametere for viktige elementer i prosjektene og - prosessene dine
- Dokumenter grenseverdiene for hvert attributt

- Visualisere prosesser og relasjoner mellom klasser, systemer og attributter ved hjelp av flytskjemaer (f.eks. i Miro, Canva, Visio)
- Automatiser datakonvertering**
 - Sjekk hvilke av dokumentene dine som ofte brukes i prosesser, som kan digitaliseres ved hjelp av OCR-biblioteker og konverteres til tabellform
 - Sjekk ut reverse engineering-verktøy for å hente ut data fra CAD (BIM)
 - Prøv å sette opp automatisk datainnhenting fra dokumenter eller formater som du ofte bruker i arbeidet ditt, til et tabellskjema
 - Konfigurer automatisk konvertering mellom ulike dataformater
- Opprett en kunnskapsbase for kategorisering**
 - Utvikle en intern eller bruke en eksisterende elementklassifisering i tråd med bransjestandarder
 - Dokumentere sammenhengen mellom ulike klassifiseringssystemer
 - Diskuter med teamet ditt om bruk av et enhetlig system for identifikasjon og entydig klassifisering av elementer
 - Begynn å bygge opp en prosess for automatisk validering av data - både de dataene dere jobber med i teamet og de som sendes til eksterne systemer

Ved å bruke disse metodene kan du forbedre datakvaliteten betydelig og forenkle den påfølgende behandlingen og transformasjonen av dataene. I de neste delene av boken vil vi se nærmere på hvordan du kan bruke allerede strukturerte og forberedte data til automatiserte beregninger, kostnadsestimering, planlegging og styring av byggeprosjekter.





V DEL

KOSTNADS- OG TIDSBEREGNINGER: INNLEMME DATA I BYGGEPROSesser

Den femte delen er viet praktiske aspekter ved bruk av data for å optimalisere kostnadsberegning og planlegging av byggeprosjekter. Den ressursbaserte metoden for kostnadsestimering og automatisering av kalkulasjonsprosesser blir analysert i detalj. Metoder for automatisert innhenting av Quantity Take-Off fra CAD (BIM)-modeller og integrering av disse med kalkulasjonssystemer blir vurdert. 4D og 5D modelleringsteknikker for tidsplanlegging og styring av byggekostnader blir utforsket, med konkrete eksempler på hvordan de kan brukes. En analyse av de utvidede informasjonslagene 6D -8D, som gir en integrert tilnærming til vurdering av bærekraft, drift og sikkerhet for eiendomsobjekter, presenteres. Metodene for å beregne karbonfotavtrykk og ESG -indikatorer for byggeprosjekter i sammenheng med moderne miljøkrav og -standarder blir vurdert i detalj. Mulighetene og begrensningene ved tradisjonelle ERP og PMIS -systemer for styring av byggeprosesser blir kritisk vurdert, med en analyse av deres innvirkning på åpenhet i prissettingen. Det spås utsikter til en overgang fra lukkede løsninger til åpne standarder og fleksible dataanalyseverktøy som kan sikre større effektivitet i byggeprosessene.

KAPITTEL 5.1.

KOSTNADSBEREGNINGER OG ESTIMATER FOR BYGGEPROSJEKTER

Grunnleggende om bygging: estimering av mengde, kostnad og tid

Blant de mange forretningsprosessene som er avgjørende for bærekraften til et selskap i byggebransjen, er prosessene for nøyaktig estimering av antall elementer, prosjektkostnader og leveringstid spesielt viktige - slik de var for tusenvis av år siden (fig. 5.1-1).

Utviklingen av skriftspråket var et resultat av en rekke faktorer, blant annet behovet for å dokumentere økonomiske transaksjoner, handel og ressursforvaltning i de tidlige samfunnene. De første juridisk betydningsfulle dokumentene, leirtavler med beregninger av materialkostnader og arbeidsavgifter, ble brukt i forbindelse med handel og bygging. Disse tavlene nedtegnet partenes forpliktelser i forbindelse med bygging av konstruksjoner, og de ble oppbevart som bevis på avtaler og penge- og handelsforbindelser.

I årtusener har tilnærmingen til estimering i stor grad vært uforandret: Beregninger ble utført manuelt, og man stoltet på kalkulasjonsingenørens erfaring og intuisjon. Men med inntoget av modulbaserte ERP-systemer og CAD-verktøy begynte den tradisjonelle tilnærmingen til estimering av mengde, kostnad og tid å endre seg raskt. Dagens digitale teknologi gjør det mulig å automatisere de viktigste beregningene av tid og kostnader, noe som gir større nøyaktighet, hastighet og åpenhet i ressursplanleggingen av byggeprosjekter.

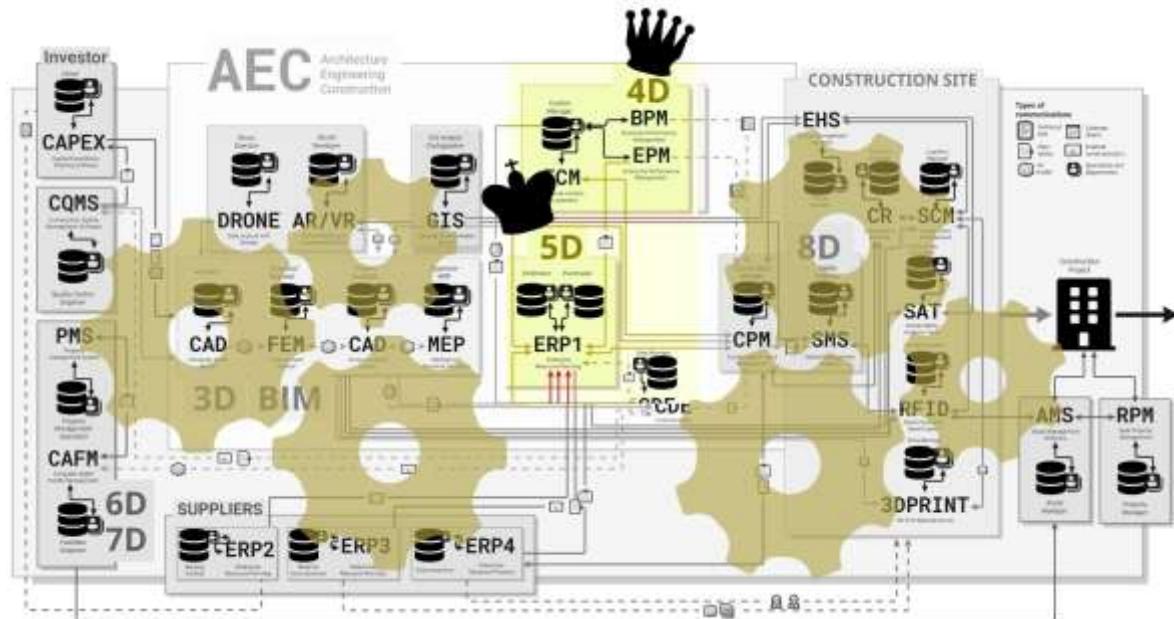


Fig. 5.1-1 Av de mange ulike systemene er verktøyene som er ansvarlige for volum-, kostnads- og tidsmålinger, de viktigste i virksomheten.

Byggefirmaer fokuserer på nøyaktige tids- og kostnadsdata. Disse er i sin tur avhengig av hvor mye

materialer og arbeidskraft som brukes, og gjennomsiktigheten påvirker lønnsomheten. Kompleksiteten i kalkulasjonsprosessen og mangelen på transparens fører imidlertid ofte til høyere prosjektkostnader, overskredne tidsfrister og til og med konkurser.

Ifølge KPMG-rapporten "Familiar Problems - New Approaches" (2023) blir bare 50% av byggeprosjektene fullført i tide, og 87% av selskapene rapporterer om økt kontroll over økonomien i kapitalprosjekter. Hovedproblemene er knyttet til mangel på kvalifisert personell og vanskeligheter med å forutse risiko [2].

Historiske data om kostnader og prosesstid samles inn under byggingen av tidligere prosjekter gjennom hele byggefirmaets levetid og legges inn i databaser i ulike systemer (ERP, PMIS BPM, EPM, osv.).

Tilgangen på historiske kostnadsdata av høy kvalitet er et stort konkurransefortrinn for en byggeorganisasjon, og påvirker direkte overlevelsesraten.

Estimerings- og kalkulasjonsavdelinger i bygg- og anleggsbedrifter er opprettet for å samle inn, lagre og oppdatere historiske data om prosjektkalkyler. Hovedfunksjonen deres er å samle og systematisere selskapets erfaring, slik at nøyaktigheten i estimeringen av omfang, tidsplan og kostnader for nye prosjekter kan bli bedre over tid. Denne tilnærmingen bidrar til å minimere feil i fremtidige beregninger basert på praksis og resultater fra allerede gjennomførte prosjekter.

Metoder for beregning av estimerte prosjektkostnader

Kalkulasjonsspesialister bruker en rekke ulike estimeringsmetoder, som alle fokuserer på en bestemt type data, tilgjengelig informasjon og prosjektets detaljnivå. De vanligste inkluderer

- **Ressursbasert metode:** Estimering av prosjektkostnadene basert på en detaljert analyse av alle ressursene som kreves, for eksempel materialer, utstyr og arbeidskraft. Denne metoden krever en detaljert liste over alle oppgavene og ressursene som kreves for å utføre hver oppgave, etterfulgt av en beregning av kostnadene. Denne metoden er svært nøyaktig og er mye brukt i kostnadsestimering.
- **Parametrisk metode:** bruker statistiske modeller for å estimere kostnader basert på prosjektparametere. Dette kan innebære at man analyserer kostnadene per måleenhet, for eksempel bygningsareal eller arbeidsomfang, og tilpasser disse verdiene til spesifikke prosjektforhold. Metoden er spesielt effektiv i de tidlige fasene, når detaljert informasjon ennå ikke er tilgjengelig.
- **Enhetsmetoden (enhetskostnadsmetoden):** Beregner den estimerte kostnaden for et prosjekt basert på en kostnad per måleenhet (f.eks. per kvadratmeter eller kubikkmeter). Dette gir en rask og enkel måte å sammenligne og analysere kostnadene for ulike prosjekter eller deler av prosjekter på.
- **Ekspertrurderinger (Delphi-metoden):** basert på meninger fra eksperter som bruker sin erfaring og kunnskap til å anslå verdien av et prosjekt. Metoden er nyttig når det ikke finnes nøyaktige grunnlagsdata, eller når prosjektet er unikt.

Det er verdt å merke seg at den parametriske metoden og ekspertrurderinger kan tilpasses maskinlæringsmodeller. Dette gjør det mulig å generere automatiske prognosenter for prosjektkostnader og tidsplaner basert på treningsseksempler. Eksempler på bruk av slike modeller er nærmere omtalt i kapittelet "Et eksempel på bruk av maskinlæring for å finne prosjektkostnader og tidslinjer" (fig. 9.3-5).

Likevel er den ressursbaserte metoden fortsatt den mest populære og mest brukte i praksis. Den gir ikke

bare en nøyaktig vurdering av de estimerte kostnadene, men gjør det også mulig å beregne varigheten av de enkelte prosessene på byggeplassen og hele prosjektet som helhet (flere detaljer i kapittelet "Byggeplaner og 4D - prosjektdata").

Ressursbasert metode kostnadsberegnung og estimering i bygg- og anleggsbransjen

Ressursbasert kalkulasjon er en metode for økonomistyring der kostnadene for et prosjekt er basert på en direkte beregning av alle involverte ressurser. I bygg- og anleggsbransjen innebærer denne tilnærmingen en detaljert analyse og evaluering av alle materialer, arbeidskraft og tekniske ressurser som kreves for å utføre arbeidet.

Den ressursbaserte metoden, gir en høy grad av transparens og nøyaktighet i budsjettplanleggingen, ettersom den fokuserer på de faktiske prisene på ressursene på estimeringstidspunktet. Dette er spesielt viktig i et ustabilt økonomisk miljø der prissvingninger kan påvirke prosjektets totalkostnad i betydelig grad.

I de følgende kapitlene skal vi se nærmere på den ressursbaserte kalkulasjonsprosessen. For å bedre forstå prinsippene i konstruksjonen, vil vi trekke en analogi med kostnadsberegningen av en middag på en restaurant. Når restaurantsjefen planlegger kvelden, lager han en liste over nødvendige produkter, tar hensyn til tilberedningstiden for hver rett og multipliserer deretter kostnadene med antall gjester. I bygg- og anleggsbransjen er prosessen den samme: For hver kategori av prosjektelementer (objekter) genereres det spesifiserte estimerer Recipes, og totalkostnaden for prosjektet bestemmes ved å summere alle kostnadene i en totalregning - det endelige estimatet per kategori.

Det viktigste og første trinnet i den ressursbaserte tilnærmingen er opprettelsen av bedriftens opprinnelige database. I den første fasen av kostnadsberegningen utarbeides det en strukturert liste over alle varer, materialer, arbeidstyper og ressurser som bedriften har til rådighet i byggeprosjektene - fra en spiker på lageret til en beskrivelse av personer med deres kvalifikasjoner og timepris. Denne informasjonen systematiseres i en felles "Byggressurs- og materialdatabase" - et tabellregister som inneholder data om navn, egenskaper, måleenheter og gjeldende priser. Det er denne databasen som blir den viktigste og primære informasjonskilden for alle etterfølgende ressursberegninger - både når det gjelder kostnader og tidsplan for arbeidene.

Database over byggeressurser: katalog over byggematerialer og byggverk

En database eller tabell over byggeressurser og -materialer - inneholder detaljert informasjon om hvert element som kan brukes i et byggeprosjekt - et produkt, en vare, et materiale eller en tjeneste, inkludert navn, beskrivelse, måleenhet og enhetskostnad, registrert i en strukturert form. I denne tabellen kan du finne alt fra ulike typer drivstoff og materialer som brukes i prosjekter, til detaljerte lister over spesialister i form av ulike kategorier med beskrivelser av timepriser (fig. 5.1-2).

Database of resources			
 1st grade potatoes 1 kg \$2,99	Sand lime bricks	 1 pcs \$1	JCB 3CX backhoe loader
 Black Angus marble beef 1 kg \$26,99	 1 h \$150	Laborer of the 1st category	
 Broccoli 1 pcs \$1,99	 1 h \$30		

Figur 5.1-2 En ressurstabell er en ingrediensliste som beskriver et materiale og en tjeneste med en enhetskostnad.

"Ressursdatabasen" kan sammenlignes med produktkatalogen i en nettbutikk, der hver vare har en detaljert beskrivelse av sine egenskaper. Dette gjør det enklere for kalkulatørene å velge de riktige ressursene (som å velge produkter når man legger dem i handlekurven) som trengs for å beregne spesifikke byggeprosesser i form av kalkyler (endelig bestilling i nettbutikken).

En ressursdatabase kan også betraktes som en liste over alle ingrediensene i en restaurantkokebok. Hvert byggemateriale, utstyr og hver tjeneste kan sammenlignes med ingrediensene som brukes i oppskriftene. "Ressursdatabasen" er en detaljert liste over alle ingrediensene - byggematerialer og tjenester, inkludert kostnadene per enhet: stykk, meter, time, liter osv.

Nye enhetselementer kan legges til i tabellen "Construction Resource Databases" på to måter - manuelt (fig. 5.1-3) eller automatisk ved å integrere med bedriftens lagerstyringssystemer eller leverandørdbaser.

Database of construction resources			
Code	Description	Unit	Unit cost
dos_p	Delivery of the mainland to the site	total	500
dos_m	Delivery to the place of installation	total	300
metprofil	Metallic profile	Kg	10.30














- Manual input

Fig. 5.1-3 Ressursdatabasen fylles ut manuelt eller henter automatisk data fra andre databaser.

Et typisk mellomstort byggefirma bruker en database som inneholder tusenvis og noen ganger titusenvis

av elementer med detaljerte beskrivelser som kan brukes i byggeprosjekter. Disse dataene brukes deretter automatisk i kontrakter prosjektdokumentasjon for å gi en nøyaktig beskrivelse av arbeidets sammensetning og prosesser.

For å holde tritt med endrede markedsforhold, for eksempel inflasjon, oppdateres attributtet "enhetskostnad" for hvert produkt (vare eller tjeneste) i ressursdatabasen (figur 5.1-3) jevnlig manuelt eller ved automatisk nedlasting av gjeldende priser fra andre systemer eller nettplattformer.

Oppdatering av enhetskostnaden for en ressurs kan gjøres månedlig, kvartalsvis eller årlig - avhengig av ressursens art, inflasjon og det eksterne økonomiske klimaet. Slike oppdateringer er nødvendige for å opprettholde nøyaktigheten i beregninger og estimater, ettersom disse grunnleggende elementene er utgangspunktet for kostnadsestimatorenes arbeid. Oppdaterte data brukes til å utarbeide estimater, budsjetter og tidsplaner som gjenspeiler de reelle markedsforholdene og reduserer risikoen for feil i senere prosjektkalkyler.

Utarbeidelse av kalkyler og kostnadsberegnning av arbeider basert på ressursgrunnlaget

Når du har fylt "Construction Resource Database" (fig. 5.1-3) med minimumsenheter, kan du begynne å lage beregninger, som beregnes for hver prosess eller arbeid på byggeplassen for bestemte måleenheter: for eksempel for en kubikkmeter betong, en kvadratmeter gipsvegg, per meter kantstein eller per vindusmontering.

For eksempel, for å bygge en murvegg på 1 m² (figur 5.1-4), basert på erfaring fra tidligere prosjekter, kreves det ca. 65 murstein (entitet "Silicate Brick") til en kostnad på 1 dollar per stykk (attributt "Cost per piece"), til sammen 65 dollar. Min erfaring tilsier også at det er nødvendig å bruke anleggsutstyr (entitet "JCB 3CX Loader") i 10 minutter for å plassere mursteinene i nærheten av arbeidsområdet. Siden det koster \$ 150 per time å leie utstyret, vil 6 minutters bruk koste ca. \$ 15. I tillegg vil det være behov for en murentreprenør i 2 timer, med en timepris på 30 dollar og en total kostnad på 60 dollar.

x0.1 1 kg \$2,99	x0.4 1 kg. \$2,99	x65 1 pcs \$1	x500 1 pcs \$1
x0.3 1 kg \$26,99	x0.4 1 kg. \$26,99	x0.1 1 h \$150	x2 1 h \$150
x0.5 1 pcs \$1,99	x0.2 1 pcs. \$1,99	x2 1 h \$30	x30 1 h \$30
+ margins and speculation	+ margins and speculation	+ margins and speculation	+ margins and speculation

Fig. 5.1-4 Kostnadskalkylene inneholder en detaljert liste over byggematerialer og tjenester som kreves for å utføre arbeidene og prosessene.

Sammensetningen av beregninger (såkalte "oppskrifter") er dannet på grunnlag av historisk erfaring akkumulert av selskapet i prosessen med å utføre et stort volum av lignende arbeid. Denne praktiske erfaringen akkumuleres vanligvis gjennom tilbakemeldinger fra byggeplassen. Arbeidslederen samler inn informasjon direkte på byggeplassen, og registrerer faktiske lønnskostnader, materialforbruk og nyanser i den teknologiske driften. I samarbeid med kalkulasjonsavdelingen blir denne informasjonen deretter iterativt forbedret: Prosessbeskrivelser blir finpasset, ressursmiksen blir justert, og kalkylene blir oppdatert for å gjenspeile faktiske data fra de siste prosjektene.

På samme måte som en oppskrift beskriver ingrediensene og mengdene som trengs for å tilberede en rett, inneholder et kalkyleark en detaljert liste over alle byggematerialer, ressurser og tjenester som kreves for å fullføre en bestemt jobb eller prosess.

Regelmessig utført arbeid gjør det mulig for arbeidere, formenn og kalkulatorer å orientere seg i den nødvendige mengden ressurser: materialer, drivstoff, arbeidstid og andre parametere som kreves for å utføre en arbeidsenhet (fig. 5.1-5). Disse dataene legges inn i kalkulasjonssystemene i form av tabeller, der hver oppgave og operasjon beskrives gjennom minimumselementene i ressursbasen (med kontinuerlig oppdaterte priser), noe som sikrer nøyaktigheten i beregningene.

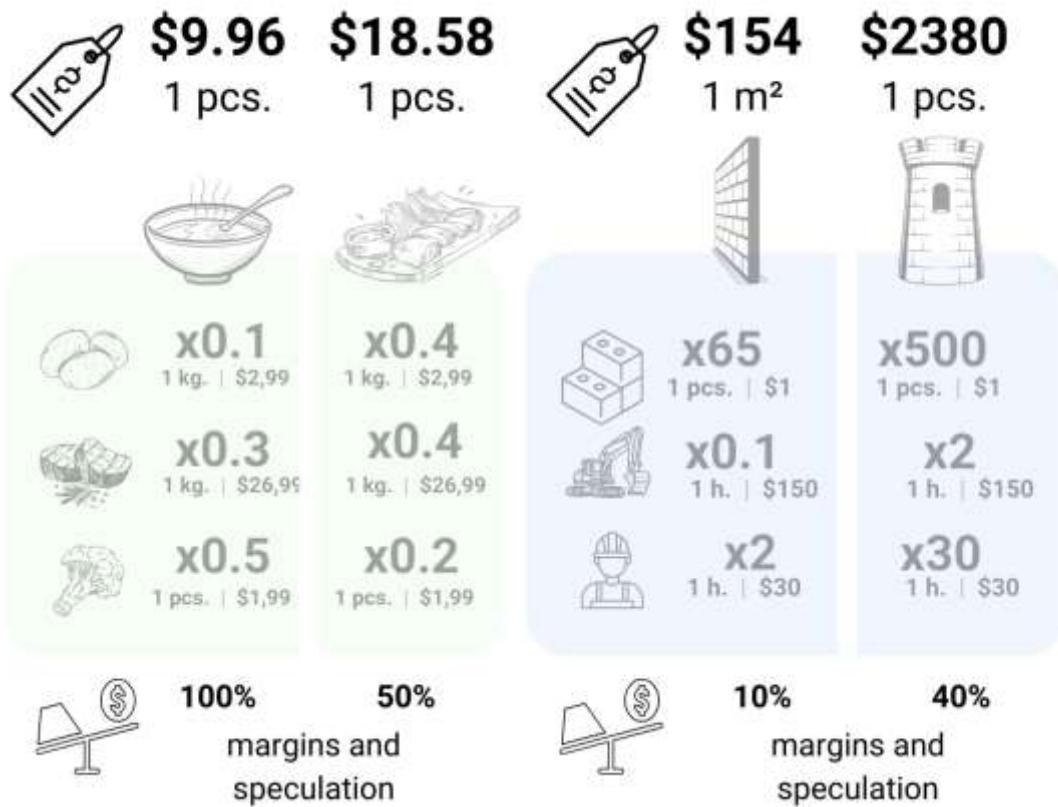


Fig. 5.1-5 Enhetspriser samles inn for hver jobb, der volumattributtet til en enhet multipliseres med mengden og tillegges en fortjenesteprosent.

For å få den totale kostnaden for hver prosess eller aktivitet (kalkyleobjekt), multipliseres kostnadsattributtet med antall og faktorer. Koeffisientene kan ta hensyn til ulike faktorer, for eksempel arbeidets kompleksitet, regionale egenskaper, inflasjonsrate, potensiell risiko (forventet overheadprosent) eller spekulasjon (ekstra profittfaktor).

Kalkulatøren, som analytiker, konverterer formannens erfaringer og anbefalinger til standardiserte kalkyler, som beskriver byggeprosesser gjennom ressursenheter i tabellform. Kalkulatørens oppgave er i hovedsak å samle inn og strukturere informasjonen byggeplassen ved hjelp av parametere og koeffisienter.

Den endelige kostnaden per arbeidsenhet (f.eks. kvadratmeter eller kubikkmeter, eller en installasjon av en enkelt enhet) inkluderer ikke bare de direkte material- og arbeidskostnadene, men også bedriftens påslag, kostnader, forsikring og andre faktorer (figur 5.1-6)

Samtidig trenger vi ikke lenger å bekymre oss for de faktiske prisene i (oppskrifts)kalkulasjoner, ettersom de reelle prisene alltid gjenspeiles i "ressursbasen" (ingredienstabellen). På kalkulasjonsnivå data fra ressursbasen lastes automatisk inn i tabellen (f.eks. etter varekode eller unik identifikator), som laster inn beskrivelsen og den faktiske prisen per enhet, som i sin tur kan lastes inn automatisk fra nettplattformer eller nettbutikker med byggematerialer. Kalkulatoren på kalkulasjonsnivå trenger bare å beskrive arbeidet

eller prosessen gjennom attributtet "ressursmengde" og tilleggsfaktorer.

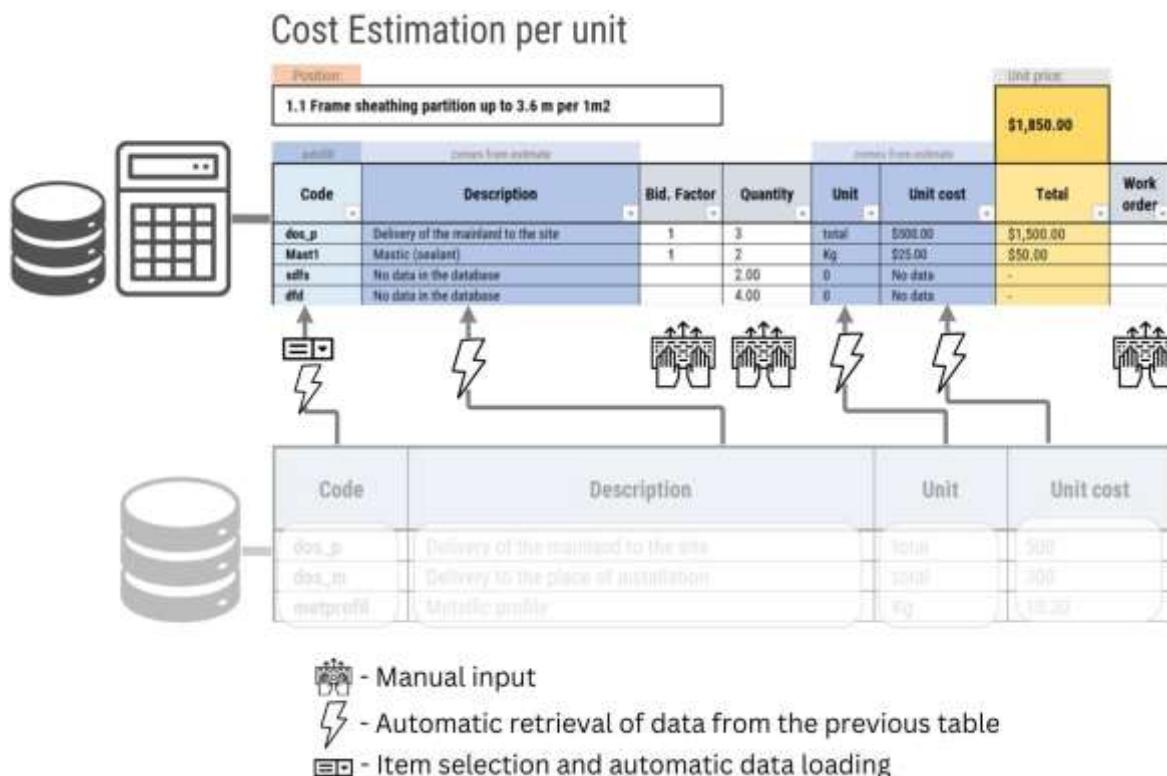


Fig. 5.1-6 I fasen for beregning av arbeidets enhetskostnad er det bare attributtene for mengden ressurser som kreves som fylles inn, alt annet lastes automatisk inn fra ressursdatabasen.

Kalkylene som opprettes, lagres i form av maler for typiske prosjekter, som er direkte knyttet til databasen over byggeressurser og -materialer. Disse malene representerer standardiserte oppskrifter for repeterende arbeidstyper for fremtidige prosjekter, noe som sikrer enhetlige kalkyler i hele selskapet.

Når kostnaden for en ressurs endres i databasen (fig. 5.1-3) - enten manuelt eller automatisk via nedlasting av gjeldende markedspriser (f.eks. i inflasjonssituasjoner) - gjenspeiles oppdateringene umiddelbart i alle tilknyttede kalkyler (fig. 5.1-6). Det betyr at det kun er ressursbasen som trenger å endres, mens kalkylemalene og estimatene forblir uendret over tid. Denne tilnærmingen sikrer stabilitet og reproducertbarhet i beregningene for eventuelle prissvingninger, som det kun tas hensyn til i en relativt enkel ressurstabell (figur 5.1-3).

For hvert nye prosjekt opprettes det en kopi av standardkalkylen, slik at du kan gjøre endringer og justere aktiviteter for å oppfylle spesifikke krav uten å endre den opprinnelige malen som er vedtatt av selskapet. Denne tilnærmingen gir fleksibilitet i tilpasningen av kalkylene: Det er mulig å ta hensyn til byggeplassens særegenheter, kundens ønsker, innføre risiko- eller lønnsomhetskoeffisienter (spekulasjon) - alt uten å bryte med bedriftens standarder. Dette hjelper selskapet med å finne en balanse mellom profittmaksimering, kundetilfredshet og opprettholdelse av konkurransesonen.

I noen land er slike kalkulasjonsmaler, som er samlet gjennom flere tiår, standardisert på nasjonalt nivå og blitt en del av de nasjonale standardene for kalkulasjonssystemer for bygg og anlegg (figur 5.1-7).

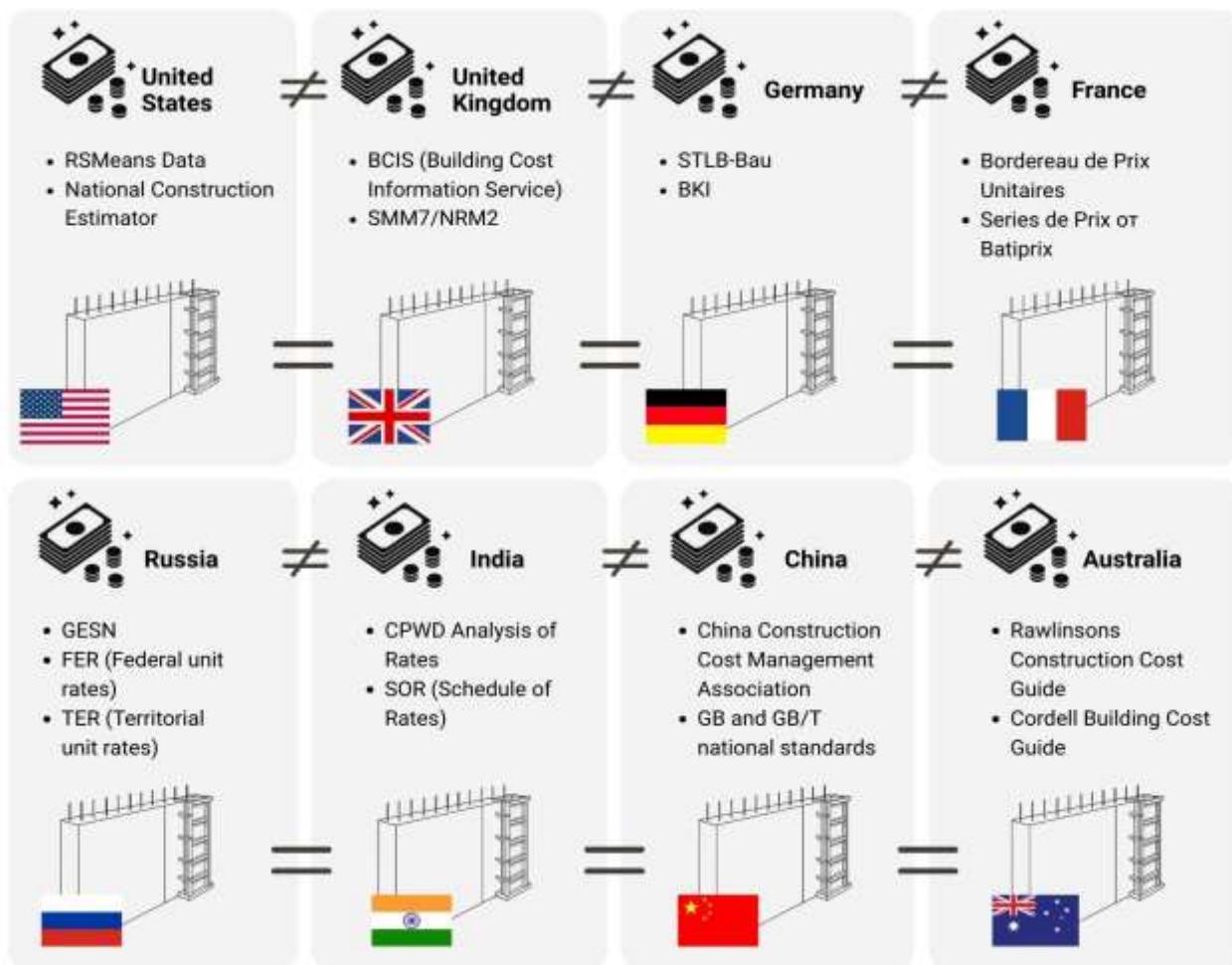


Fig. 5.1-7 Ulike land rundt om i verden har sine egne kalkulasjonsregler med egne (preskriptive) kompendier og standarder for bygg- og anleggsarbeid for kalkulasjon av det samme elementet.

Slike standardiserte ressursbaser (figur 5.1-7) er obligatoriske for alle markedsaktører, spesielt for offentlig finansierte prosjekter. En slik standardisering sikrer transparens, sammenlignbarhet og rettferdighet i prisdannelsen og kontraktsforpliktelser for kunden.

Sluttkalkulasjon av prosjekter: fra estimator til budsjetter

Statlige og bransjespesifikke kalkulasjonsstandarder spiller forskjellige roller i byggepraksis i ulike land. Mens noen land krever streng overholdelse av én enkelt standard, har de fleste utviklede økonomier en mer fleksibel tilnærming. I markedsøkonomier fungerer statlige byggestandarder vanligvis bare som en grunnlinje. Byggefirmaene tilpasser disse standardene til sine egne driftsmodeller eller reviderer dem fullstendig, og supplerer dem med sine egne tilpassede faktorer. Disse justeringene gjenspeiler bedriftens erfaring, ressursforvaltningseffektivitet og ofte faktorer der for eksempel et selskaps

spekulasjonsfortjeneste kan være tatt med i betraktingen.

Resultatet er at konkurransenivået, etterspørlsen i markedet, målmarginer og til og med forholdet til spesifikke kunder kan føre til betydelige avvik fra standardiserte normer. Denne praksisen gir fleksibilitet i markedet, men gjør det også vanskelig å sammenligne tilbud fra ulike entreprenører på en transparent måte, noe som introduserer et element av spekulativ prising i byggebransjen på dette stadiet av kalkulasjonsprosessen.

Når beregningsmalene for de enkelte aktivitetene og prosessene er utarbeidet - eller, som oftest, bare kopiert fra offentlige standardkalkyler (fig. 5.1-7) med koeffisienter lagt til for å gjenspeile "særegenheterne" til en bestemt bedrift - er det siste trinnet å multiplisere kostnadene for hver post med den tilsvarende egenskapen for arbeidsomfanget eller prosessene i det nye prosjektet.

Når man skal beregne totalkostnaden for et nytt byggeprosjekt, er det viktigste steget å summere kostnadene for alle kalkylepostene, multiplisert med volumet av disse arbeidspostene i prosjektet.

For å lage totalkostnaden for prosjektet, i vårt forenklede eksempel, starter vi med å beregne kostnaden for å bygge én kvadratmeter vegg og multipliserer kostnaden for beregningen (f.eks. arbeidet "1m² standard installasjon av veggelementer") med det totale antallet kvadratmeter veger i prosjektet (f.eks. attributtet "Areal" eller "Mengde" (fig. 5.1-8) for en enhet av typen "Veggelementer" fra CAD i prosjektet eller kalkulasjonene til arbeidslederen).

På samme måte beregner vi kostnaden for alle elementene i prosjektet (fig. 5.1-8): Vi tar kostnaden per arbeidsenhet og multipliserer den med volumet av et bestemt element eller en gruppe elementer i prosjektet. Kalkulatøren trenger bare å legge inn antallet av disse elementene, aktivitetene eller prosessene i prosjektet i form av volum eller mengde. På denne måten kan man automatisk generere et komplett byggekalkyle.

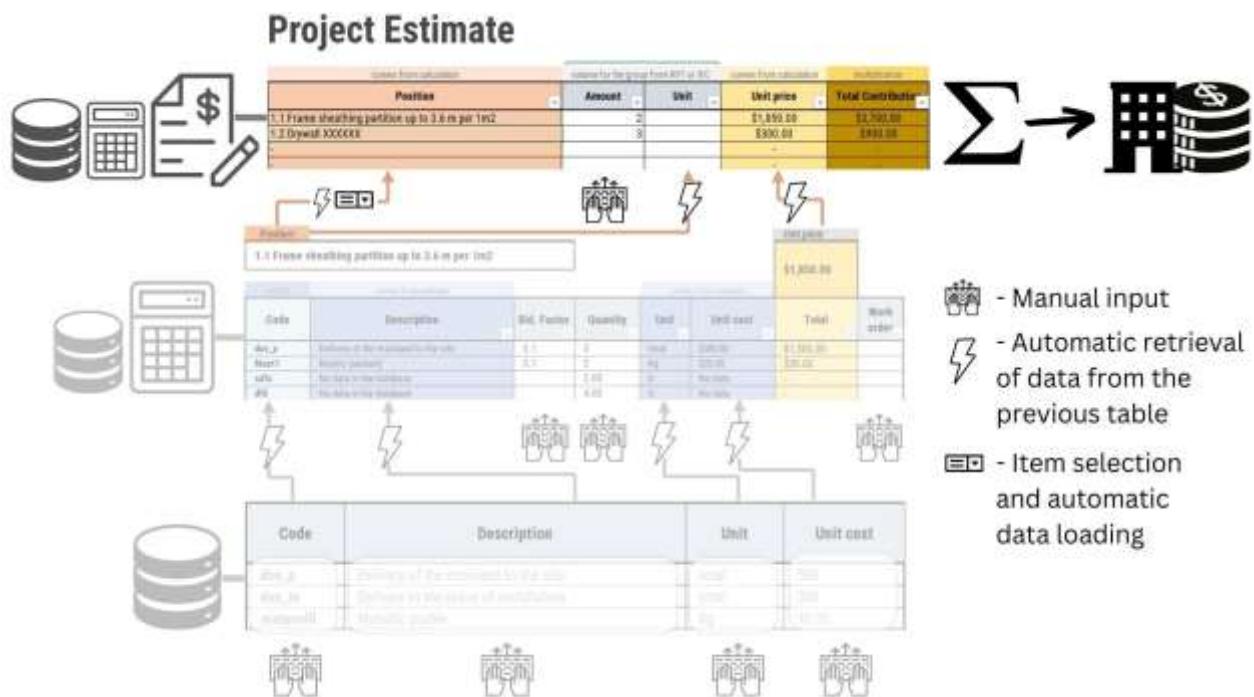


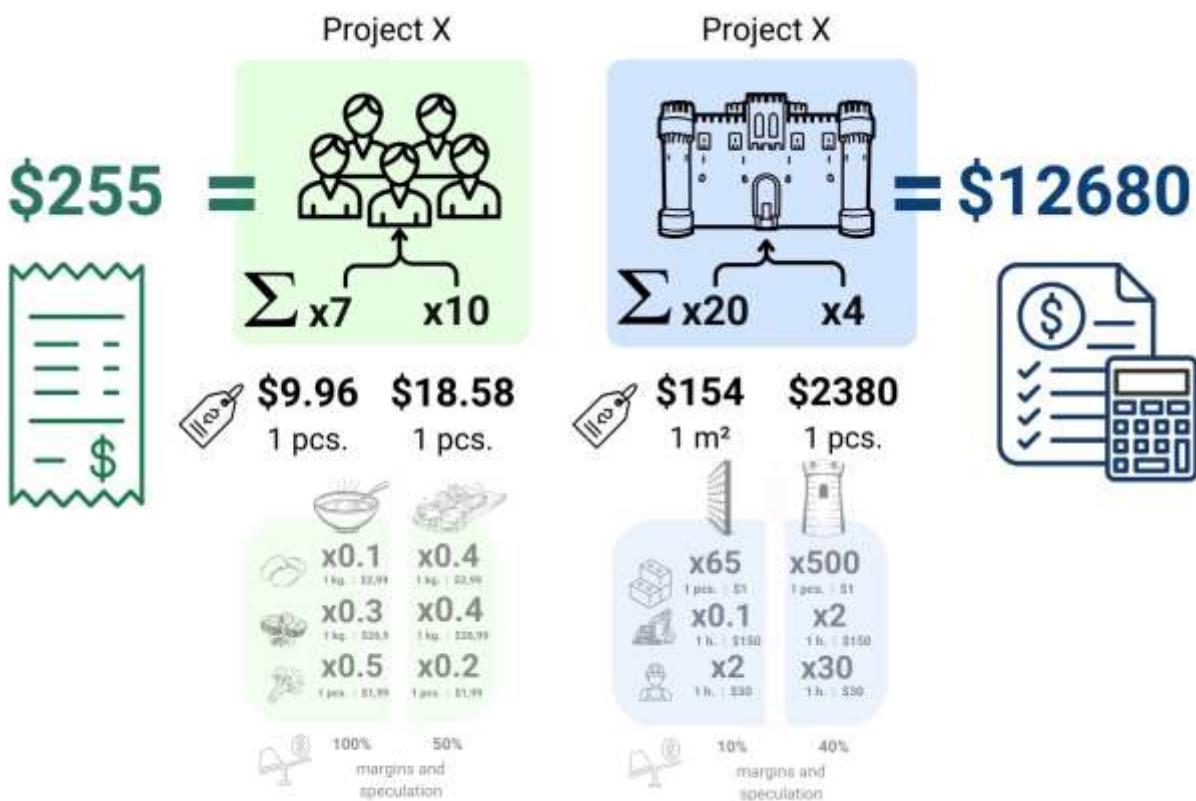
Fig. 5.1-8 Når vi lager et estimat, legger vi bare inn arbeidsomfanget.

På samme måte som for kalkyler, laster vi på dette nivået automatisk opp ferdig kalkulerte poster (fra kalkylemalen eller nye som er kopiert fra malen og redigert), som automatisk tar med seg den aktuelle kostnaden per arbeidseenhet (som oppdateres automatisk fra ressursdatabasen (fig. 5.1-8 nederste tabell)). Hvis det skjer endringer i ressursdatabasen eller kalkyletabellene, vil dataene i kalkylen automatisk bli oppdatert for den aktuelle dagen, uten at det er nødvendig å endre selve kalkylen eller estimatet.

I restaurantsammenheng beregnes den endelige kostnaden for et arrangement på en lignende måte og tilsvarer den endelige kostnaden for hele middagen, der kostnaden for hver rett multiplisert med antall gjester summerer seg til den totale kostnaden for sjekken (figur 5.1-9). Og akkurat som i byggebransjen kan oppskriftene for matlaging på en restaurant ligge fast i flere tiår. I motsetning til prisene, der råvarekostnadene kan endre seg hver time.

På samme måte som en restauranteier multipliserer kostnaden for hvert måltid med antall porsjoner og personer for å finne totalkostnaden for arrangementet, legger kalkulasjonslederen sammen kostnadene for alle prosjektkomponentene for å komme frem til et fullstendig byggeoverslag.

Dermed bestemmes den endelige kostnaden for hver aktivitet i prosjektet (fig. 5.1-9), som multiplisert med attributtvolumet til enheten som tilsvarer denne aktiviteten - gir kostnadene for grupper av aktiviteter, hvorfra den endelige kostnaden for hele prosjektet oppnås.



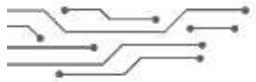
Figur 5.1-9 Det endelige estimatet beregnes ved å summere hvert elements arbeidskostnadsattributt med dets omfangsattributt.

Den totale prosjektkostnaden (figur 5.1-8) gir et økonomisk bilde av prosjektet, slik at kunder, investorer eller finansieringsorganisasjoner kan forstå det totale budsjettet og de økonomiske ressursene som kreves for å levere prosjektet på en gitt dag, tatt i betraktning gjeldende priser.

Og hvis prosessene for å samle ressursbaser, beregninger og estimer (prosessoppskrifter) allerede er utarbeidet, halvautomatisert og finpusset i titusenvis av år og registrert på statsnivå, er automatisk innhenting av kvalitativ informasjon om volum og mengde elementer for den siste fasen av det endelige estimatet - i dag fortsatt en flaskehals i prosessene for alle beregninger av kostnads- og tidsattributter for prosjektet, og generelt det samlede budsjettet for prosjektet.

I tusenvis av år har den tradisjonelle metoden for å beregne volumer vært manuelle metoder for å måle volum og mengde ved hjelp av flate tegninger. Med den digitale tidsalderen har bedrifter oppdaget at volum- og mengdeinformasjon nå kan hentes ut automatisk fra geometriske data i CAD-modeller, noe som revolusjonerer tusen år gamle metoder for innhenting av kvantitative data.

Moderne tilnæringer til prosesskalkulasjon og estimering innebærer automatisk utvinning av volumetriske og kvantitative attributter fra CAD-databaser, som kan lastes opp og kobles til kalkulasjonsprosessen for å få oppdaterte prosjektgruppevolumer på ethvert trinn fra design til drift.



KAPITTEL 5.2.

MENGDEBEREGNING OG AUTOMATISK OPPRETTING AV ESTIMATER OG TIDSPLANER

Fra 3D til 4D og 5D: bruk av volumetriske og kvantitative parametere

Med kalkulasjonstabellene med de beskrevne prosessene gjennom ressurser (fig. 5.1-8) i hånden, er neste trinn å automatisk hente inn volum- eller mengdeparametere for en gruppe elementer som er nødvendige for beregningene og for det endelige estimatet.

Volumetriske egenskaper for prosjektelementer - f.eks. veger eller plater - kan automatisk hentes ut fra CAD-databaser. Parametriske objekter som er opprettet i CAD-programmer, konverteres ved hjelp av en geometrikjerne til numeriske verdier for lengde, bredde, areal, volum og andre parametere. Prosessen med å fremskaffe volumer basert på 3D-geometri vil bli diskutert mer detaljert i den neste, sjette delen (fig. 6.3-3), som er dedikert til arbeid med DAK (BIM). I tillegg til volumer kan man også få antall lignende elementer fra CAD-modelldatabasen ved å filtrere og gruppere objekter etter kategorier og egenskaper. Disse parameterne, som gjør det mulig å gruppere, blir grunnlaget for å knytte sammen prosjektelementer gjennom ressursberegninger med kalkyler, endelige estimer og budsjett for hele prosjektet.

Datamodellen som hentes ut fra 3D-modellen (CAD), utvides med nye parameterlag, kalt 4D og 5D. I de nye lagene med enhetsattributter, 4D (tid) og 5D (kostnad), brukes geometriske 3D-data som kilde til attributtverdier for enhetsvolum.

- **4D** er et parameterinformasjonslag som legger til informasjon om varigheten av byggeoperasjoner til 3D-parametrene for elementer. Disse dataene er avgjørende for planlegging av tidsplaner og styring av prosjektets tidslinjer
- **5D** er det neste nivået av utvidelse av datamodellen, der elementene suppleres med kostnadskarakteristikker. Dette tilfører et økonomisk aspekt til den geometriske informasjonen: kostnader for materialer, arbeider og utstyr, noe som muliggjør budsjettberegnninger, lønnsomhetsanalyser og kostnadsstyring under byggeprosessen.

Kostnads- og 3D, 4D og 5D -attributtdata for grupper av prosjektenheter beskrives på samme måte som beregninger i modulære ERP-, PIMS- og -systemer (eller Excel -lignende verktøy) og brukes til automatisk kostnadsberegnung og budsjettplanlegging av både individuelle grupper og hele prosjektbudsjettet.

5D-attributter og innhenting av attributtvolum fra CAD

Ved utarbeidelse av det endelige estimatet for et byggeprosjekt, som vi har diskutert i tidligere kapitler (figur 5.1-8), blir volumattributtene for hver kategori av prosjektelementer enten samlet inn manuelt eller hentet fra volumattributtspesifikasjonene som leveres av CAD-programvaren.

Den tradisjonelle manuelle metoden for beregning av mengder innebærer at formann og kalkulator analyserer tegninger som i tusenvis av år har vært presentert som linjer på papir og de siste 30 årene i digitale formater som PDF (PLT) eller DWG. Med utgangspunkt i sin yrkesfaring måler de opp arbeids- og materialmengder, ofte med linjal og vinkelmåler. Denne metoden krever betydelig innsats og tid, samt spesiell oppmerksomhet på detaljer.

Det kan ta alt fra noen få dager til flere måneder å fastsette omfangsegenskaper på denne måten, avhengig av prosjektets omfang. Siden alle målinger og beregninger gjøres manuelt, er det i tillegg en risiko for menneskelige feil som kan føre til unøyaktige data, noe som igjen kan føre til feil i estimeringen av prosjektets tid og kostnader, noe som hele selskapet vil bli holdt ansvarlig for.

Moderne metoder basert på bruk av CAD-databaser forenkler beregningen av volumer betraktelig. I DAK-modeller inneholder geometrien til elementene allerede volumattributter som kan beregnes automatisk (via geometrikjernen (fig. 6.3-3)) og presenteres eller eksporteres i tabellform.

I et slikt scenario ber kalkulasjonsavdelingen CAD-designeren om data om prosjektlelementenes mengde- og volumegenskaper. Disse dataene eksporteres som regneark eller integreres direkte i kalkulasjonsdatabaser - enten det er Excel, ERP eller PMIS - systemer. Denne prosessen starter ofte ikke med en formell forespørrelse, men med en kort dialog mellom kunden (initiativtakeren) og arkitekten/kalkulatøren fra bygge- eller designfirmaet. Nedenfor er et forenklet eksempel som viser hvordan en strukturert tabell for automatiske beregninger (QTO) dannes ut fra dagligdags kommunikasjon:

- ⌚ Kunde - "Jeg ønsker å legge til en etasje til i bygningen, i samme konfigurasjon som første etasje"
- ⌚ Arkitekt (CAD) - "Legger til en tredje etasje, konfigurasjonen er den samme som i andre etasje". Og etter denne meldingen sendes en ny CAD-versjon av prosjektet til kalkulatoren.
- ⌚ Estimator utfører automatisk gruppering og beregning (ERP, PMIS, Excel) - "Jeg kjører prosjektet gjennom et Excel-regneark med QTO-regler (ERP, PMIS), henter volumene etter kategori for den nye etasjen genererer et estimat".

Resultatet er at tekstdialogen blir omgjort til en tabellstruktur med grupperingsregler:

Element	Kategori	Gulv
Overlapping	OST_Floors	3
Spalte	OST_StructuralColumns	3

Element	Kategori	Gulv
trappeoppgang	OST_Trappene	3

Etter prosessen med automatisk gruppering av CAD-modellen fra designeren i henhold til QTO-reglene i kalkulatoren og automatisk multiplikasjon av volumer med ressursberegninger (fig. 5.1-8) får vi følgende resultater, som sendes til kunden:

Element	Volum	Gulv	Pris per enhet.	Total kostnad
Overlapping	420 m ²	3	150 €/m ²	63 000 €
Spalte	4 stk.	3	2450 €/stikk.	9 800 €
trappeoppgang	2 stk.	3	4 300 €/stikk.	8 600 €
TOTALT:	-	-	-	81 400 €

- ☒ Kunden - "Takk, det er ganske mye, vi må kutte ned noen rom". Og syklusen gjentar seg mange ganger.

Dette scenariet kan gjentas mange ganger, spesielt i godkjenningsfasen, der kunden forventer umiddelbar tilbakemelding. I praksis kan slike prosesser imidlertid trekke ut i dager eller til og med uker. Takket være innføringen av automatiske grupperings- og beregningsregler kan aktiviteter som tidligere tok lang tid, i dag gjennomføres på få minutter. Automatisert innhenting av mengder ved hjelp av grupperingsregler gjør ikke bare beregningene og estimatene raskere, men ved å minimere den menneskelige faktoren reduseres også sannsynligheten for feil, noe som gir en transparent og nøyaktig vurdering av prosjektkostnadene.

Hvis kravene til estimeringsavdelingen opprinnelig ble tatt i betraktning ved opprettelsen av 3D-modellen i CAD-systemet (noe som fortsatt er sjeldent i praksis), og navnene, identifikatorene til elementgrupper og deres klassifiseringsattributter er angitt i form av parametere som sammenfaller med strukturene til estimeringsgrupper og -klasser, kan volumetriske attributter automatisk overføres til estimeringssystemene uten ytterligere transformasjoner.

Automatisk uthenting av volumattributter fra CAD i form av spesifikasjonstabeller gjør det mulig å raskt få oppdaterte data om kostnadene for de enkelte arbeidene og prosjektet som helhet (fig. 5.2-1). Ved å oppdatere kun CAD-filen med prosjektvolumene i kalkulasjonsprosessen eller i kalkulasjonssystemet, kan selskapet raskt beregne kalkylen på nytt ved å ta hensyn til de siste endringene, noe som sikrer høy nøyaktighet og konsistens i alle påfølgende kalkyler.

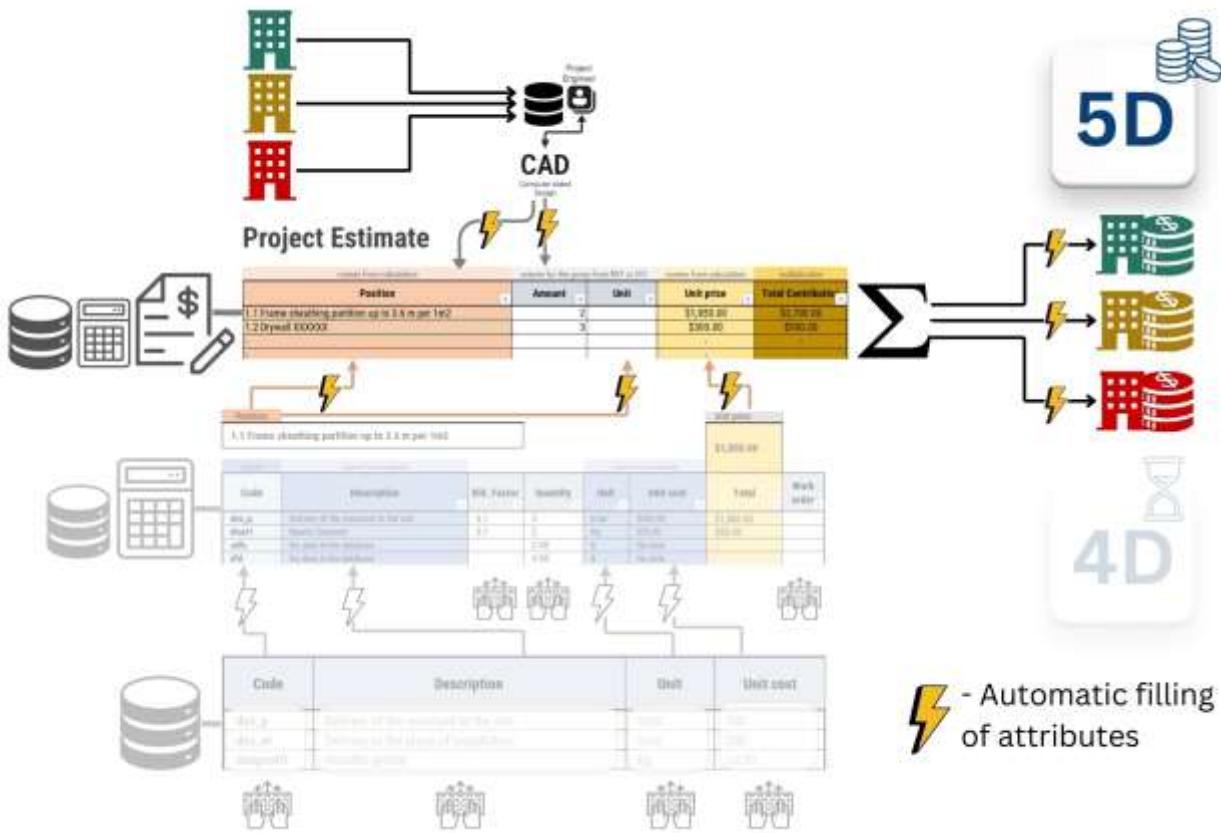


Fig. 5.2-1 Volumattributter fra CAD-tabeller eller databaser legges automatisk inn i kalkylen, slik at du umiddelbart kan beregne den totale prosjektkostnaden.

Med den økende kompleksiteten i kapitalprosjekter blir det å beregne hele budsjettet og analysere de totale prosjektkostnadene under et slikt scenario (figur 5.2-1) - et viktig verktøy for å kunne ta informerte beslutninger.

Ifølge Accentures studie Creating More Value through Capital Projects (2024) [20] integrerer ledende selskaper dataanalyse aktivt i digitale initiativer, og bruker historisk informasjon til å forutsi og optimalisere resultater. Forskning viser at flere eiere og operatører bruker stordataanalyser for å forutsi markedstrender og vurdere kommersiell levedyktighet før prosjekteringen starter. Dette gjøres ved å analysere datavarehus fra en eksisterende portefølje av prosjekter. I tillegg implementerer 79 prosent av eier-operatørene "robuste" prediktive analyser for å vurdere prosjektresultater og støtte operasjonelle beslutninger i sanntid.

Moderne, effektiv styring av byggeprosjekter er uløselig knyttet til behandling og analyse av store mengder informasjon på alle stadier av prosjekteringen og de prosessene som går forut for prosjekteringen. Bruken av datavarehus, ressursberegninger, prediktive modeller og maskinlæring gjør det ikke bare mulig å minimere risikoen i beregningene, men også å ta strategiske beslutninger om prosjektfinansiering på et tidlig stadium av prosjekteringen. Vi kommer til å snakke mer om datavarehus og prediktive modeller som utfyller kalkylene i den niende delen av boken.

Automatisk innhenting av volumetriske parametere for elementer fra CAD-prosjekter, som er nødvendige for utarbeidelse av estimater, utføres ved hjelp av grupperingsverktøyet QTO (Quantity Take-Off). QTO-

verktøyene grupperer alle prosjektobjekter etter spesielle elementidentifikatorer eller elementattributparametere, ved hjelp av spesifikasjoner og tabeller som er opprettet i CAD-databasen.

QTO Quantity Take-Off: gruppering av prosjektdata etter attributter

QTO (Quantity Take-Off) i bygg- og anleggsbransjen er en prosess som går ut på å hente ut mengdene av elementer som kreves for å realisere et prosjekt. I praksis er QTO ofte en semi-manuell prosess som involverer innsamling av data fra ulike kilder: PDF-dokumenter, DWG-tegninger og digitale CAD-modeller.

Når man arbeider med data hentet fra CAD-databaser, utføres QTO-prosessen som en sekvens av filtrerings-, sorterings-, grupperings- og aggregeringsoperasjoner. Modellelementer velges ut i henhold til parametrene for klasser, kategorier og typer, og deretter oppsummeres de kvantitative attributtene deres - som volum, areal, lengde eller mengde - i henhold til beregningslogikken (fig. 5.2-2).

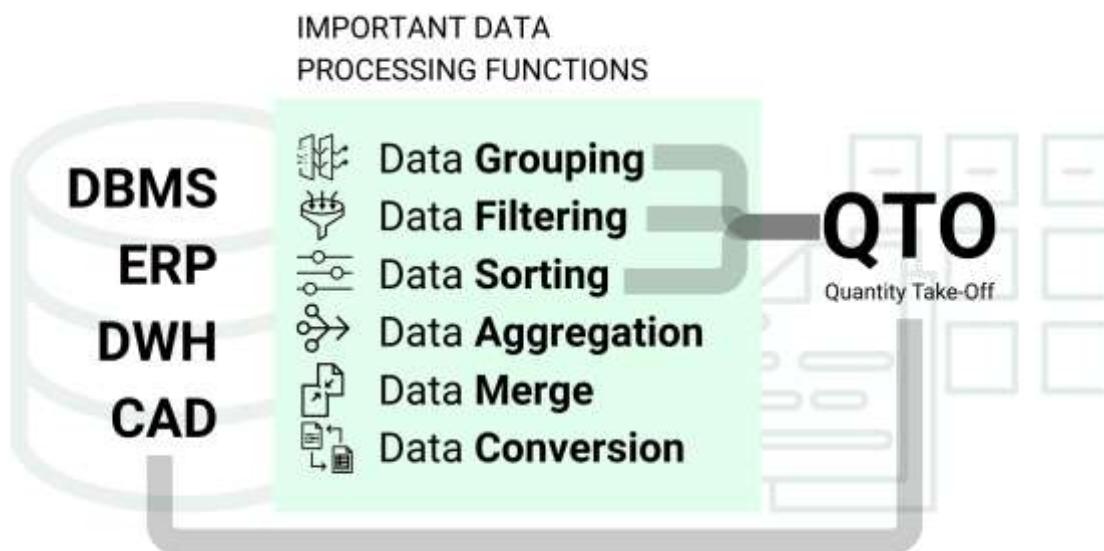


Fig. 5.2-2 Gruppering og filtrering av data er de mest populære funksjonene som brukes i databaser og datavarehus.

QTO-prosessen (filtrering og gruppering) gjør det mulig å systematisere data, lage spesifikasjoner og forberede inngangsinformasjon for beregning av estimatorer, innkjøp og arbeidsplaner. Grunnlaget for QTO er klassifisering av elementer etter type målte attributter. For hvert element eller gruppe av elementer velges den tilsvarende kvantitative måleparameteren. For eksempel

- **Lengdeattributt** (kantstein - i meter)
- **Arealattributt** (gipsplatearbeid - i kvadratmeter)
- **Volumattributt** (betongarbeider - i kubikkmeter)
- **Mengdeattributt** (vinduer - per stykk)

I tillegg til de volumetriske egenskapene som genereres matematisk på grunnlag av geometrien, brukes ofte overskridelsesfaktorer (fig. 5.2-12, f.eks. 1,1 for å ta høyde for 10 prosent for logistikk og montering) i beregningene etter QTO-grupperingen - korreksjonsverdier som tar hensyn til svinn, montering, lagring eller

transport. Dette gjør det mulig å forutsi det faktiske materialforbruket mer nøyaktig og å unngå både under- og overlager på byggeplassen.

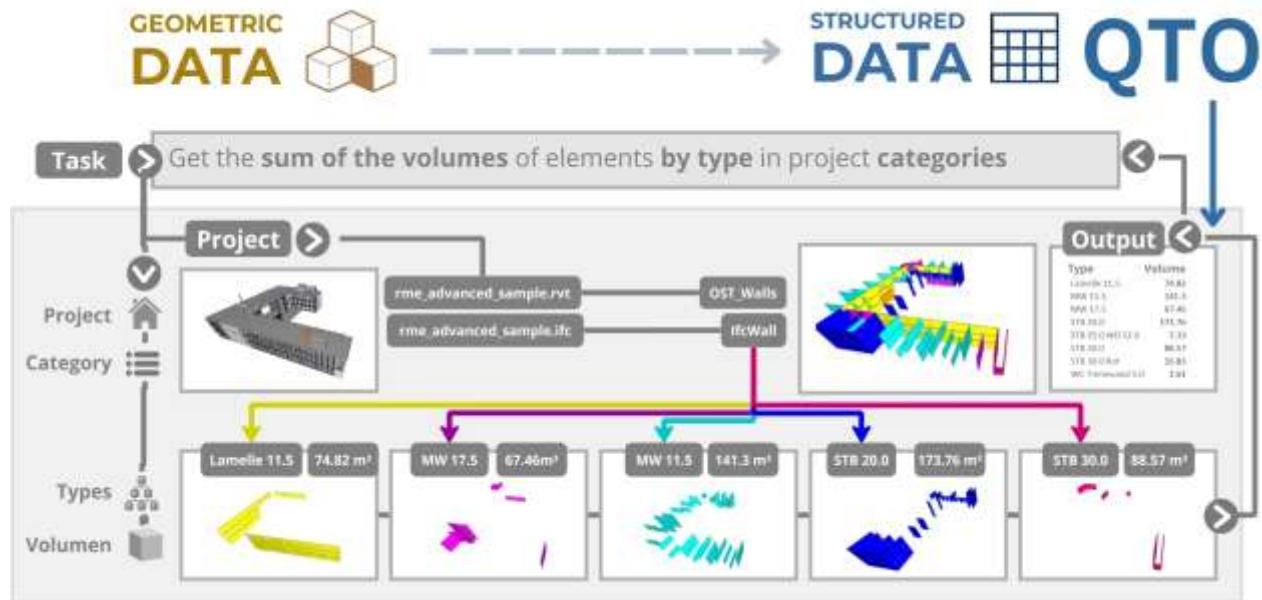
En automatisert mengdeuttaksprosess (QTO) er avgjørende for å produsere nøyaktige beregninger og estimater, redusere menneskelige feil i volumspesifikasjons prosesser og forhindre over- eller underbestillinger av materialer.

Som et eksempel på QTO-prosessen kan vi se på et vanlig tilfelle der det er nødvendig å vise en tabellspesifikasjon av volumer etter elementtyper for en bestemt kategori eller klasse av elementer fra CAD-databasen. La oss gruppere alle prosjektelementer etter type fra CAD-prosjektveggkategorien og oppsummere volumattributtene for hver type for å presentere resultatet som en QTO-volumtabell (fig. 5.2-3).

I et eksempel på et typisk CAD-prosjekt (fig. 5.2-3) er alle veggkategorieelementer i CAD-databasen gruppert etter veggtypes, f.eks. "Lamelle 11.5", "MW 11.5" og "STB 20.0", og har veldefinerte volumattributter representert i metriske kuber.

Målet for prosjektlederen, som befinner seg i grensesnittet mellom konstruktører og beregningsspesialister, er å få en automatisert tabell over volumer per elementtype i den valgte kategorien. Ikke bare for et spesifikt prosjekt, men også i en universell form som kan brukes på andre prosjekter med lignende modellstruktur. På denne måten blir tilnærmingen skalerbar, og data kan gjenbrukes uten dobbeltarbeid.

Den tiden er forbi da erfarne konstruktører og kalkulatører bevæpnet seg med linjal og nøye målte opp hver eneste linje på papir eller PDF -planer - en tradisjon som ikke har endret seg de siste årtusenene. Med utviklingen av 3D -modellering, der geometrien til hvert enkelt element nå er direkte knyttet til automatisk beregnede volumetriske attributter, har prosessen med å bestemme volumer og QTO -mengder blitt automatisert.



Figur 5.2-3 For å hente ut QTO volum- og mengdeattributter fra et prosjekt må man gruppere og filtrere prosjektelementer.

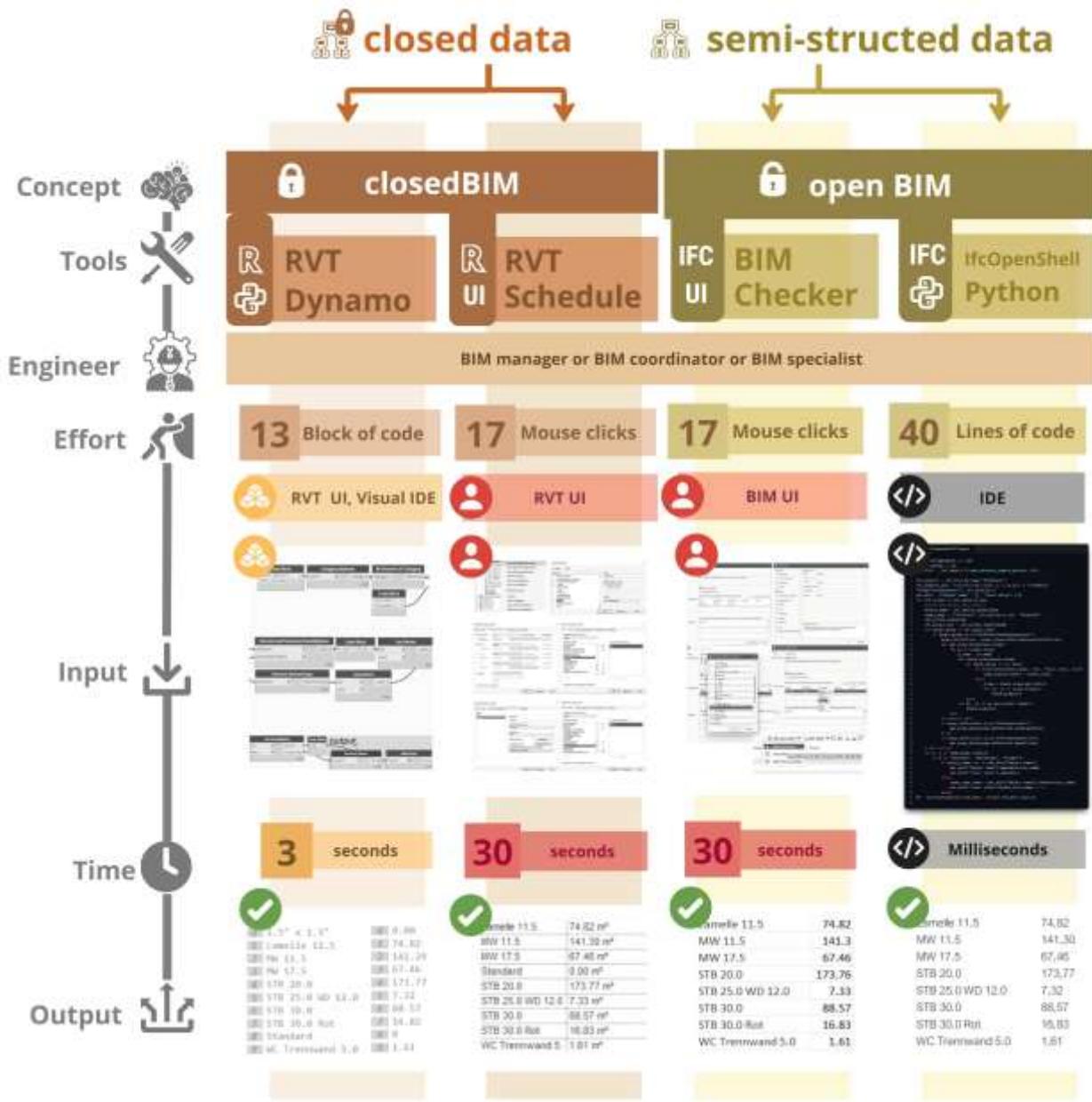
I vårt eksempel er oppgaven å "velge en kategori av vegg i et prosjekt, gruppere alle elementer etter type og presentere informasjon om volumattributter i et strukturert, tabellarisk format", slik at denne tabellen kan brukes av dusinvis av andre spesialister til beregninger av kostnader, logistikk, arbeidsplaner og andre forretningsaker (fig. 6.1-3).

På grunn av CAD-dataenes lukkede natur kan ikke alle spesialister i dag bruke direkte tilgang til CAD-databasen (årsakene til og løsningene på tilgangsproblemene er beskrevet i bokens sjette del). Derfor må mange ty til spesialiserte BIM-verktøy basert på konseptene åpen BIM og lukket BIM [63]. Når man arbeider med spesialiserte BIM-verktøy eller direkte i CAD-programvaremiljøet, kan tabellen med QTO-resultater (Quantity Take-Off) genereres på forskjellige måter - avhengig av om man bruker et manuelt grensesnitt eller automatiserer programvaren.

For eksempel er det nok å bruke brukergrensesnittet til en CAD (BIM)-programvare for å utføre ca. 17 handlinger (klikk på knapper) for å få en ferdig volumtabell (fig. 5.2-4). Brukeren må imidlertid ha en god forståelse av modellstrukturen og funksjonene til CAD (BIM)-programvaren.

Hvis automatiseringen skjer via programkode eller via plugin-moduler og API-verktøy i CAD-programmer, reduseres antallet manuelle trinn for å få tak i volumtabellene, men det må skrives mellom 40 og 150 linjer med kode, avhengig av hvilket bibliotek eller verktøy som brukes:

- **IfcOpSh (åpen BIM)** eller **Dynamo IronPython (lukket BIM)** - gjør det mulig å hente en QTO-tabell fra et CAD-format eller CAD-program på bare ~40 linjer med kode.
- **IFC_js (åpen BIM)** - krever omtrent 150 kodelinjer for å trekke ut omfattende attributter fra IFC-modellen.
- **Grensesnitt for CAD-verktøy (BIM)** - gjør at du kan oppnå samme resultat manuelt, med 17 museklikk.



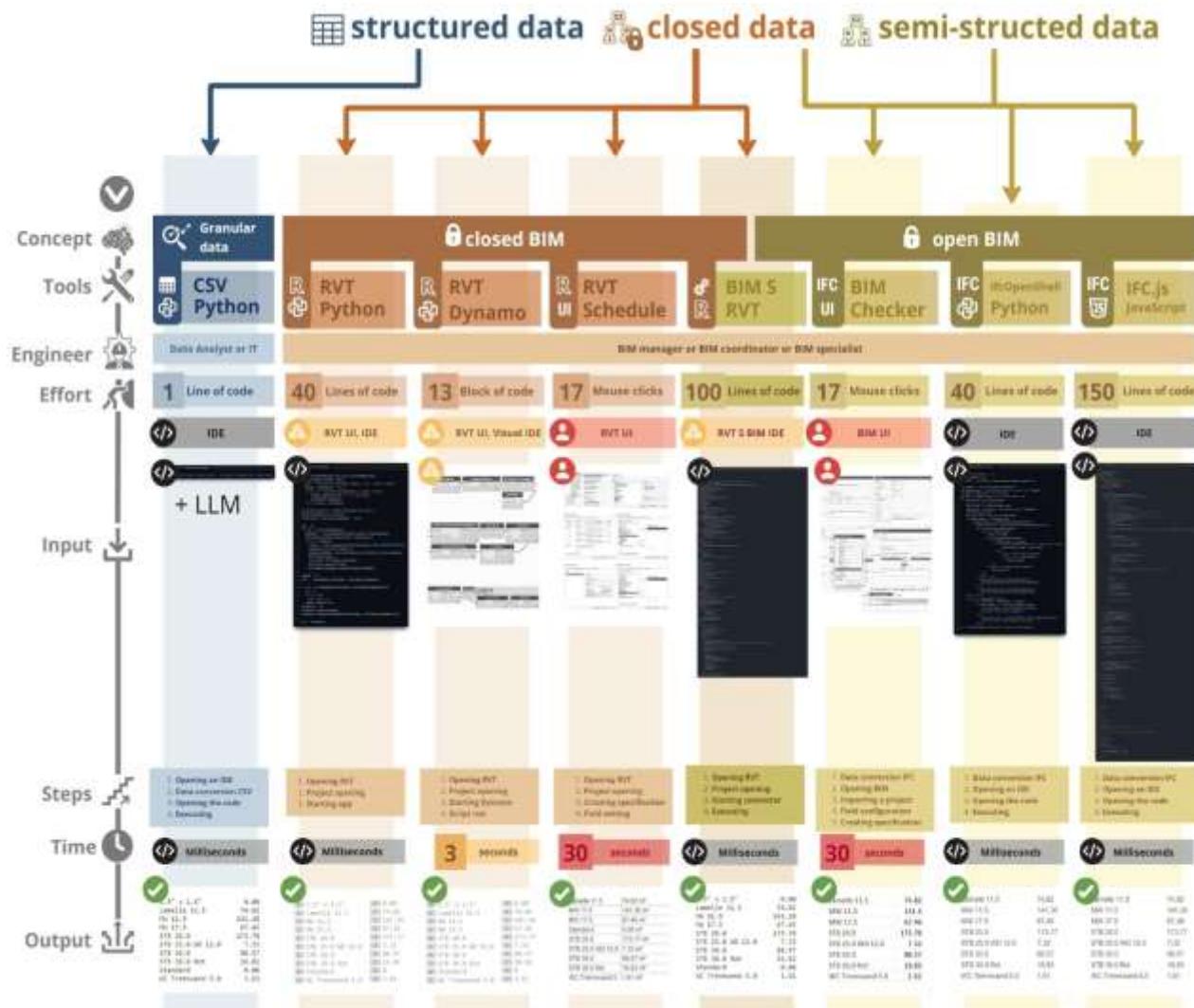
Figur 5.2-4 CAD (BIM) designere og forvaltere, bruker 40 til 150 kodelinjer eller et dusin tastetrykk for å lage QTO-tabeller

Resultatet er det samme - en strukturert tabell med volumattributter for en gruppe elementer. Den eneste forskjellen er arbeidskostnadene og den tekniske ekspertisen som kreves av brukeren (fig. 5.2-4). Moderne verktøy gjør QTO-prosessen betydelig raskere og reduserer sannsynligheten for feil i forhold til manuell innsamling av volumer. De gjør det mulig å hente data direkte fra prosjektmodellen, slik at det ikke lenger er nødvendig å beregne volumene manuelt på nytt fra tegninger, slik man gjorde tidligere.

Uansett hvilken metode som brukes - enten det er åpen BIM eller lukket BIM - er det mulig å få en identisk QTO - tabell med prosjektelementvolumer (fig. 5.2-4). Når man arbeider med prosjektdata i CAD - (BIM-)konsepter, er brukerne imidlertid avhengige av spesialiserte verktøy og API-er som leveres av

leverandørene (fig. 3.2-13). Dette skaper flere lag av avhengighet og krever at man lærer seg unike dataskjemaer, samtidig som det begrenser den direkte tilgangen til dataene.

På grunn av CAD-dataenes lukkede natur kompliserer innhenting av QTO-tabeller og andre parametere automatisering av beregninger og integrering med eksterne systemer. Ved å bruke verktøy for direkte tilgang til databaser og overføre CAD -data fra prosjektet ved hjelp av reverse engineering-verktøy til et åpent strukturert datarammeformat (fig. 4.1-13), kan en identisk QTO -tabell oppnås med bare én kodelinje (fig. 5.2-5 - variant med granulære data).



Figur 5.2-5 Ulike verktøy gir samme resultat i form av attributtabeller over prosjektenheter, men med ulike arbeidskostnader.

Ved bruk av åpne, strukturerte data fra CAD -prosjekter, som nevnt i kapittelet "Konvertering av CAD (BIM)-data til strukturert form", forenkles grupperingsprosessen, QTO, betraktelig.

Tilnæringer basert på bruk av åpne strukturerte data eller direkte tilgang til CAD-modelldatabaser er fri for de markedsføringsbegrensningene som er knyttet til akronymet BIM. De baserer seg på velprøvde verktøy som lenge har vært i bruk i andre bransjer (fig. 7.3-10 ETL prosess).

Ifølge McKinsey-studien "Open Data: Unleash Innovation and Productivity with Streaming Information" [102] fra 2013 kan bruk av åpne data skape muligheter for besparelser på mellom 30 og 50 milliarder dollar per år innen design, prosjektering, innkjøp og bygging av kraftproduksjonsanlegg. Dette tilsvarer en besparelse på 15 prosent i byggekostnadene.

Arbeid med åpne strukturerte (granulære) data forenkler informasjonsinnehenting og -behandling, reduserer avhengigheten av spesialiserte BIM-plattformer og åpner for automatisering uten behov for å bruke proprietære systemer eller parametriske og komplekse datamodeller fra CAD-formater.

QTO-automatisering ved hjelp av LLM og strukturerte data

Oversettelse av ustukturerte data til en strukturert form forbedrer effektiviteten i ulike prosesser betydelig: Det forenkler databehandlingen (fig. 4.1-1, fig. 4.1-2) og fremskynder valideringsprosessen ved å gjøre kravene klare og gjennomsiktige, som vi allerede har diskutert i de foregående kapitlene. På samme måte forenkler oversettelsen av CAD-data (BIM) til en strukturert, åpen form (fig. 4.1-12, fig. 4.1-13) attributtgrupperingsprosessen og QTO-prosessen.

Attributtabellen i QTO har en strukturert form, så når vi bruker strukturerte CAD -data, arbeider vi med én enkelt datamodell (figur 5.2-5), noe som eliminerer behovet for å konvertere og oversette prosjektdatamodeller og grupperingsregler til en fellesnevner. Dermed kan vi gruppere data etter ett eller flere attributter med bare én kodelinje. I åpen BIM og lukket BIM, der data lagres i halvstrukturerte, parametriske eller lukkede formater, krever behandlingen derimot flere titalls eller hundrevis av kodelinjer, og bruk av API for å samhandle med geometri- og attributtinformasjon.

- ☞ Eksempel på gruppering av et QTO-strukturert prosjekt etter ett attributt. Tekstforespørrelse i et LLM chatterom (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller et annet):

Jeg har CAD -prosjekt som DataFrame - vennligst filtrer prosjektdataene for å få elementer med "Type" -parameteren som bare inneholder "Type 1 -verdien. ↴

- LLMs svar vil sannsynligvis være i form av Python-kode ved hjelp av Pandas :



Fig. 5.2-6 En kodelinje skrevet med LLM , gjør det mulig å gruppere et helt CAD -prosjekt etter "Type"-attributtet og få den ønskede gruppen av elementer.

Takket være den enkle strukturen i 2D DataFrame trenger vi ikke å forklare LLM -skjemaet og datamodellen, noe som forkorter tolkningsfasen og gjør det raskere å lage endelige løsninger. Tidligere var det nødvendig å lære seg programmeringsspråk for å skrive selv enkel kode, men nå lar moderne språkmodeller (LLM) oss automatisk oversette prosesslogikk til kode når vi arbeider med strukturerte data ved hjelp av tekstspøringer.

LLM automatisering og språkmodeller kan fullstendig eliminere behovet for fagfolk som arbeider med CAD (BIM) datagruppering og -behandling fra å måtte lære seg programmeringsspråk eller BIM-verktøy ved å tilby muligheten til å løse problemer ved hjelp av tekstspøringer.

Den samme spørringen - gruppering av alle prosjektelementer fra kategorien "vegger" og beregning av volumer for hver type (fig. 5.2-5) - som i et CAD-miljø (BIM) krever 17 klikk i grensesnittet eller skriving av 40 kodelinjer, ser i åpne databehandlingsverktøy (f.eks. SQL eller Pandas) ut som en enkel og intuitiv spørring:

- Med en enkelt linje i Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

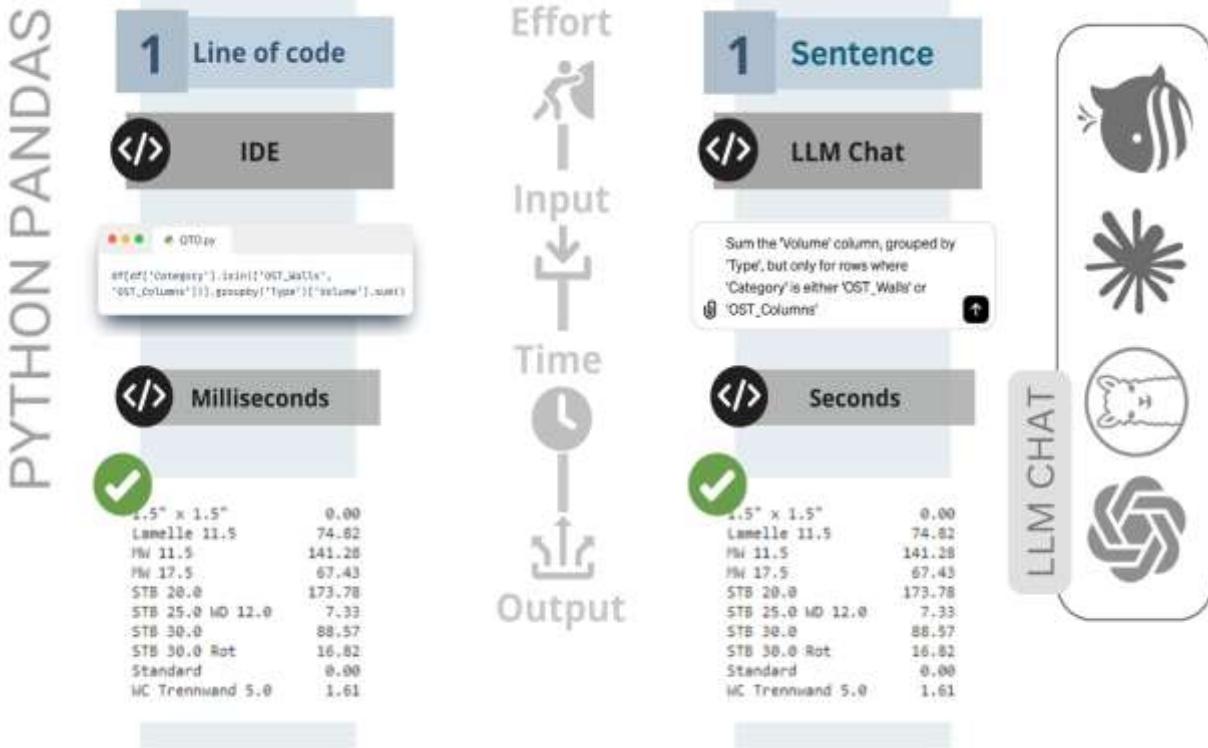
Kodeavkoding: ta fra df (DataFrame) elementene, som attributt-kolonnen "Category", har verdier "OST_Walls", gruppere alle oppnådde elementer etter attributt-kolonnen "Type" og summer opp for den oppnådde gruppen av elementer attributtet "Volume".

- Gruppering av et strukturert prosjekt hentet fra CAD ved hjelp av SQL:

```
SELECT Type, SUM(Volum) AS TotalVolume
FRA elementer
WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- Ved hjelp av LLM kan vi skrive en grupperingsforespørsel til prosjektdatabasen som en enkel tekstreferanse - en prompt (fig. 5.2-7):

For prosjektdatarammen grupperer du elementene etter parameteren "Type", men bare for elementer der parameteren "Kategori" er "OST_Walls" eller "OST_Columns", og summerer kolonneparameteren "Volum" for den resulterende ↴



Figur 5.2-7 Ved hjelp av SQL, Pandas og LLM er det nå mulig å automatisere databehandling med noen få linjer kode og tekstspørreng.

Å innhente QTO fra CAD-data ved hjelp av LLM-verktøy (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), endrer de tradisjonelle metodene for å hente ut attributtinformasjon, kvantitative og volumetriske data for enkeltobjekter og grupper av objekter på en dramatisk måte.

Nå kan selv prosjektledere, kalkulasjons- eller logistikkspesialister som ikke har inngående kunnskap om design og ikke har spesialisert CAD-programvare - (BIM-), og som har tilgang til CAD-databasen, få det totale volumet av elementer i kategorien vegg eller andre objekter i løpet av sekunder, ganske enkelt ved å skrive eller diktere en spørring.

I tekstspøringer (fig. 5.2-8) behandler LLM-agenten i modellen brukerens forespørsel om bruke en bestemt funksjon på en eller flere parametere - kolonner i tabellen. Resultatet er at brukeren i kommunikasjon med LLM mottar enten en ny kolonneparameter med nye verdier eller en spesifikk verdi etter gruppering

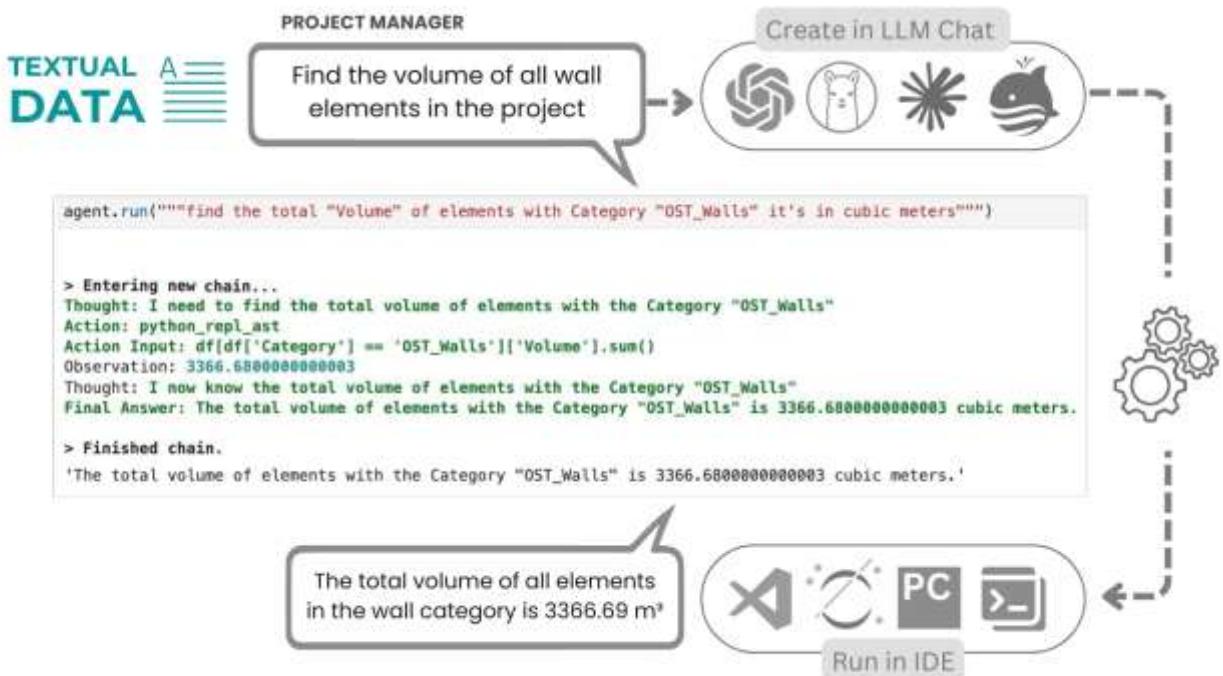


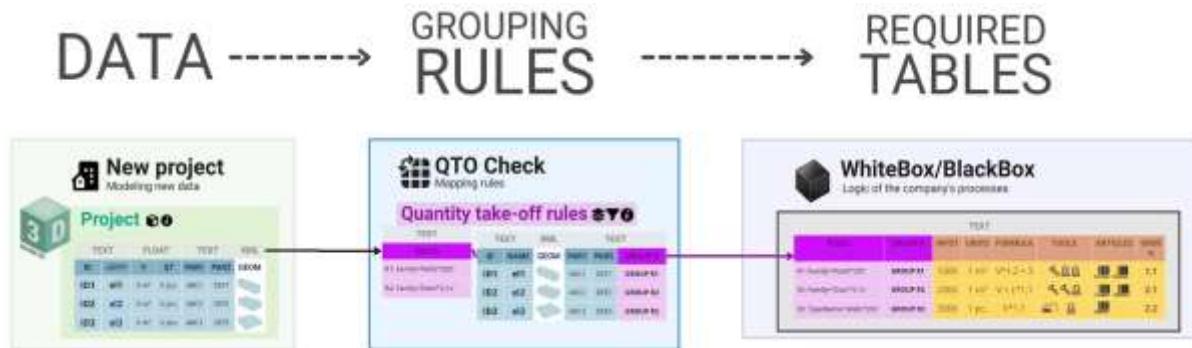
Fig. 5.2-8 LLM modellen, når den arbeider med strukturerte data, forstår ut fra konteksten i en tekstforespørsel hvilken gruppering og hvilke attributter brukeren spør om.

Hvis det er nødvendig å hente ut mengder for bare én gruppe elementer, er det tilstrekkelig å utføre en enkel QTO-spørring (figur 5.2-7) på DAK-modelldataene. Men når man skal beregne et budsjett eller estimat for et helt prosjekt som består av mange grupper av elementer, er det ofte nødvendig å hente ut mengder for alle typer elementer (klasser0 , der hver kategori av elementer behandles separat - med gruppering etter relevante attributter.

I praksis brukes individuelle grupperings- og beregningsregler for ulike typer objekter. For eksempel grupperes vinduer vanligvis etter etasjer eller soner (grupperingsparameter - attributt Nivå, Rom), og vegger - etter materiale eller konstruksjonstype (parameter Materiale, Type). For å automatisere grupperingsprosessen beskrives slike regler på forhånd i form av grupperingsregeltabeller. Disse tabellene fungerer som konfigurasjonsmaler som definerer hvilke attributter som skal brukes i beregningene for hver gruppe av elementer i prosjektet.

QTO beregning av hele prosjektet ved hjelp av grupperegler fra et Excel-regnark

I reelle byggeprosjekter er det ofte nødvendig å utføre aggregering etter flere attributter samtidig innenfor en gruppe elementer. Når du for eksempel arbeider med kategorien "Vinduer" (der Category-attributtet inneholder verdier som OST_Windows eller IfcWindows), kan elementene gruppere ikke bare etter type - for eksempel etter verdien i Type Name- eller Type-feltet - men også etter ytterligere egenskaper, for eksempel varmeledningsevnenvårt som er angitt i det tilsvarende attributtet. Denne flerdimensjonale grupperingen gir mer nøyaktige resultater for en bestemt gruppe. Ved beregning av vegg- eller gulvkategorier kan vilkårlige kombinasjoner av attributter - for eksempel materiale, nivå, gulv, brannmotstand og andre parametere - brukes som filtre eller grupperingskriterier (figur 5.2-9).



Figur 5.2-9 For hver gruppe eller kategori av enheter i et prosjekt finnes det en egen grupperingsformel som består av ett eller flere kriterier.

Prosesssen med å definere slike grupperingsregler ligner prosessen med å opprette datakrav som er beskrevet i kapittelet "Opprette krav og kvalitetssjekke data" (figur 4.4-5), der vi diskuterte arbeidet med datamodeller i detalj. Slike grupperings- og beregningsregler sikrer nøyaktigheten og relevansen av resultatene for automatisk å beregne de totale attributtene for mengden eller volumet til en kategori av enheter, med hensyn til alle de nødvendige forholdene som må tas hensyn til i beregninger og kalkulasjoner.

- 💡 Følgende kodeeksempel filtrerer prosjekttabellen slik at det resulterende datasettet bare inneholder enheter der attributtkolonnen "Category" inneholder verdiene "OST_Windows" eller "IfcWindows", samtidig som attributtkolonnen "Type" inneholder verdien "Type 1":

Jeg har et DataFrame -prosjekt - filtrer dataene slik at bare elementer som har attributtet "Category" som inneholder verdiene "OST_Windows" eller "IfcWindows", og samtidig inneholder Type-attributtet verdien "Type 1, forblir i datasettet ↴

💡 LLMs svar :

```
df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

Figur 5.2-10 Med en enkelt kodelinje, som ligner på Excel-formelen , kan du gruppere alle prosjektets enheter etter flere attributter.

Den resulterende koden (fig. 5.2-10) etter oversettelse av CAD-data i strukturerte åpne formater (fig. 4.1-13) kan kjøres i en av de populære IDE-ene (integrerte utviklingsmiljøene) vi nevnte ovenfor i frakoblet modus: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære nettbaserte verktøy: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

💡 For å hente ut prosjekt-enhetene i QTO DataFrame-skjemaet under kategorien "Windows" kun med en bestemt varmeledningseverdi, kan vi bruke følgende spørring til LLM:

Jeg har et DataFrame -prosjekt - filtrer dataene slik at bare poster med "Category" som inneholder "OST_Windows"- eller "IfcWindows"-verdier blir værende i datasettet, og samtidig skal kolonnen ThermalConductivity ha en verdi på 0,✉

💡 LLMs svar :

```
df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

Figur 5.2-11 Det ekstremt enkle spørrespråket Pandas Python gjør det mulig å kjøre QTOer for et hvilket som helst antall prosjekter samtidig.

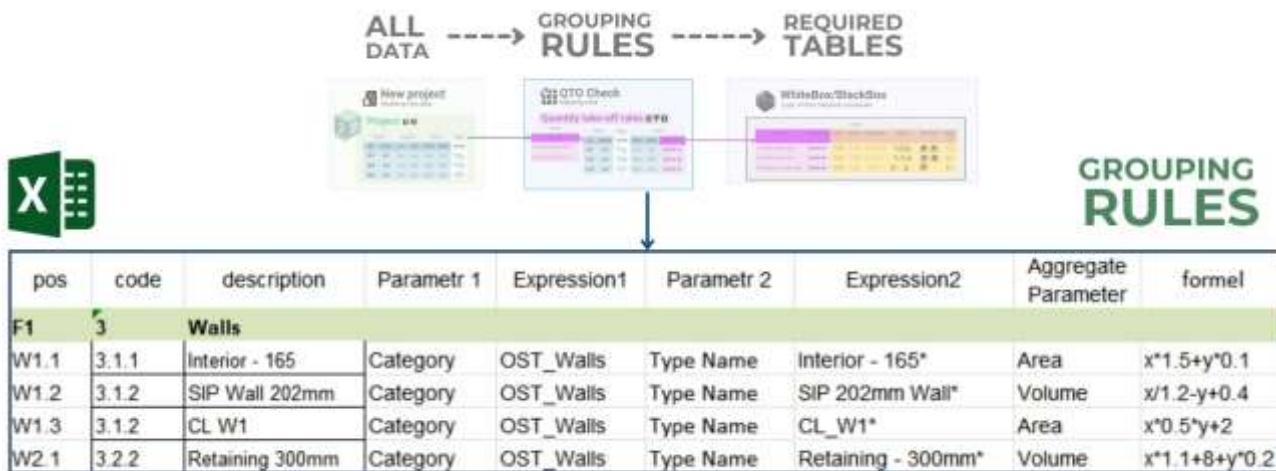
I svaret som mottas fra LLM (fig. 5.2-11), brukes den logiske betingelsen "&" til å kombinere to kriterier:

varmeledningsevne og tilhørighet til en av de to kategoriene. Metoden "isin" kontrollerer om verdien i attributt-kolonnen "Category" finnes i den angitte listen.

I prosjekter med et stort antall elementgrupper, med ulik grupperingslogikk - for hver kategori av prosjektelementer (f.eks. vinduer, dører, plater) bør det etableres individuelle grupperingsregler, som kan omfatte tilleggskoeffisienter eller formler for beregning av totalattributt. Disse formlene (figur 5.2-12 attributt "formel", f.eks. x-verdien av mengde og y-volumet av gruppe) og koeffisientene tar hensyn til de unike egenskapene til hver gruppe, for eksempel:

- % tillegg til materialvolumet for å ta høyde for overskridelser
- fast ekstra mengde materiale
- justeringer knyttet til mulige risikoer og beregningsfeil i form av formler

Når filtrerings- og grupperingsreglene er formulert i form av parameterformler for hver enkelt varekategori, kan de lagres som en linje-for-linje-tabell - for eksempel i Excel-format (figur 5.2-12). Ved å lagre disse reglene i en strukturert form kan prosessen med å hente ut, filtrere og gruppere prosjektdata automatiseres fullt ut. I stedet for å skrive mange separate spørninger manuelt, leser systemet ganske enkelt parametertabellen og bruker de aktuelle reglene på modellen (den overordnede prosjektdatarammen (fig. 4.1-13)), og genererer endelige QTO-tabeller for hver kategori av prosjektelementer.



Figur 5.2-12 QTO-attributtgruppering Tabell fastsetter regler for gruppering av prosjektelementer, slik at man får et nøyaktig totalt antall og omfang for hver kategori.

De innsamlede reglene lar deg gruppere hele prosjektet og utføre alle nødvendige beregninger, inkludert justeringer av volumatributter. Som et resultat blir volumene brukt til det "virkelige volumet", som brukes til beregninger og beregninger, og ikke de som opprinnelig var på designstadiet i CAD-modellen.

I prosessen med å automatisk opprette volumtabeller for QTO for hele prosjektet, må programmet gå gjennom alle kategoriene i grupperingsregeltabellen, ta grupperingsattributtene, gruppere alle prosjektelementene etter dem og aggregere volumattributtet for den gruppen, i tillegg til å multiplisere det med en foredlingsfaktor eller koeffisient.

La oss be LLM om å skrive kode for oss for en slik løsning, der koden må laste inn to tabeller - en tabell med grupperingsregler (Figur 5.2-12) og en tabell med data fra selve prosjektet (Figur 4.1-13), og deretter bruke grupperingsreglene, gruppere elementene i henhold til de gitte reglene, beregne de aggregerte verdiene og lagre resultatene i en ny Excel-fil.

- Send en tekstforespørsel til LLM chat:

Jeg trenger kode for å lese prosjektdata fra filen "basic_sample_project.xlsx" og deretter
regler fra "Grouping_rules_QTO.xlsx" for å gruppere alle data etter "Parameter 1" og
"Parameter 2", aggregere "Aggregate Parameter", filtrere etter "Expression2", utføre
beregninger fra "Formel1" og lagre QTO-tabellen til "QTQ_table2.xlsx". ↵

- LLMs svar :



```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 J.apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d*,\d*)!')[0]).fillna(0))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     l1[row["Aggregate Parameter"]].agg(['sum', 'count'])
17     GROUP = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
18
19     # Perform calculation defined in the rules
20     group['After Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
21     return group.reset_index()
22
23 # Apply the function to each row in the grouping rules dataframe and concatenate the results
24 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
25
26 # Save the processed data to an Excel file using the OpenPyXL engine
27 dfgroups.to_excel('QTQ_table2.xlsx', engine='openpyxl', index=False)

```

Figur 5.2-13 Python-kode Pandas behandler designdata fra et CAD-prosjekt og grupperer dem i henhold til regler som er definert i Excel .

Det endelige resultatet av kodekjøringen (fig. 5.2-13) vil være en entitetsgruppetabell som ikke bare inneholder de generaliserte volumattributtene fra den opprinnelige CAD- (BIM-) modellen, men også et nytt reelt volumattributt som tar hensyn til alle kravene for korrekt opprettelse av beregninger og estimatorer (eksempel fig. 5.2-14).

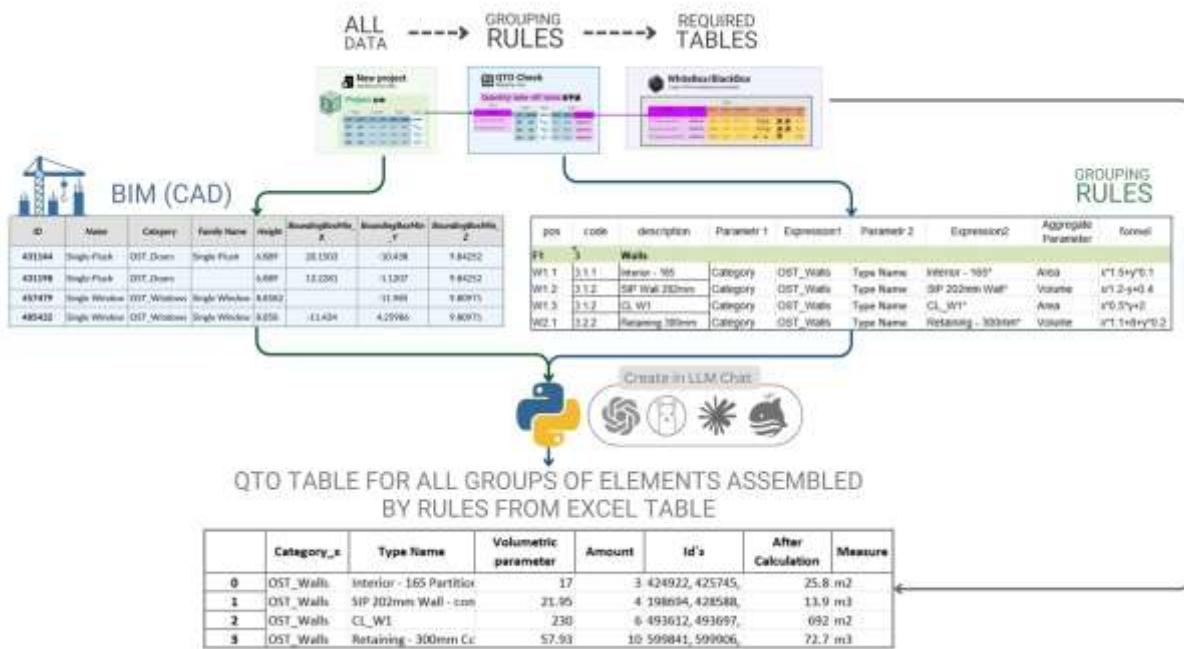
**QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE**



	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,	25.8 m2	
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,	13.9 m3	
2	OST_Walls	CL_W1		230	6 493612, 493697,	692 m2	
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,	72.7 m3	

Figur 5.2-14 Attributtet "After Calculation" legges til i oppsummeringstabellen etter at koden som automatisk beregner det faktiske volumet, er kjørt.

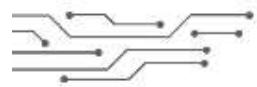
Den resulterende koden (fig. 5.2-13) kan kjøres i en av de populære IDE-ene (som vi nevnte ovenfor) og brukes på et hvilket som helst antall eksisterende eller nye innkommende prosjekter (RVT, IFC, DWG, NWS, DGN etc.), enten det dreier seg om noen få prosjekter eller kanskje hundrevis av prosjekter i forskjellige formater i en strukturert form (fig. 5.2-15).



Figur 5.2-15 Den automatiske grupperingsprosessen for konstruksjonsdata kobler BIM-data (CAD) til QTO-tabeller via regler fra et Excel-regneark.

Den tilpassede og parametriserte volumetriske datainnsamlingsprosessen (figur 5.2-15) muliggjør helautomatisk innsamling av data om kvantitative attributter og volumer av prosjektelementer for videre arbeid med dem, inkludert kostnadsestimering, logistikk, arbeidsplaner og beregning av karbonfotavtrykk og andre analytiske oppgaver.

Etter å ha lært oss verktøy som gjør det enkelt å organisere og gruppere grupper av prosjektelementer i henhold til bestemte attributter, er vi nå klare til å integrere grupperte og filtrerte prosjekter med ulike bedriftskalkyler og forretningsscenarioer.



KAPITTEL 5.3.

4D, 6D OG BEREGNING AV KARBONDIOKSIDUTSLIPP CO₂

4D-modell: Integrering av tid i konstruksjonsestimater

I tillegg til kostnadsberegning er en av de viktigste anvendelsene av prosjekteringsdata i bygg- og anleggsbransjen fastsettelse av tidsparametere - både for de enkelte byggeoperasjonene og for hele prosjektet. Den ressursbaserte kalkulasjonsmetoden og den tilhørende kalkulasjonsdatabasen, som ble diskutert i detalj i forrige kapittel "Kalkyler og estimater for byggeprosjekter", brukes ofte som grunnlag for automatisert tidsberegning og oppretting av en tidsplan for utførelsen av arbeidet.

Den ressursbaserte tilnærmingen tar ikke bare hensyn til materialkostnader, men også tidsressurser. Ved kalkulasjon kan hver prosess tilordnes et arbeidsordreattributt (figur 5.3-1 - Arbeidsordreparameter) og den tid og kostnad som er forbundet med utførelsen av denne prosessen. Disse parameterne er spesielt viktige for å beskrive operasjoner som ikke har en fast markedspris og ikke kan kjøpes direkte - for eksempel bruk av anleggsmaskiner, ansettelse av arbeidere eller logistiske prosesser (som vanligvis uttrykkes i timer). I slike tilfeller fastsettes ikke kostnaden av innkjøpsavdelingen, men direkte av det utførende selskapet basert på interne normer eller produksjonsrater (figur 5.3-1).

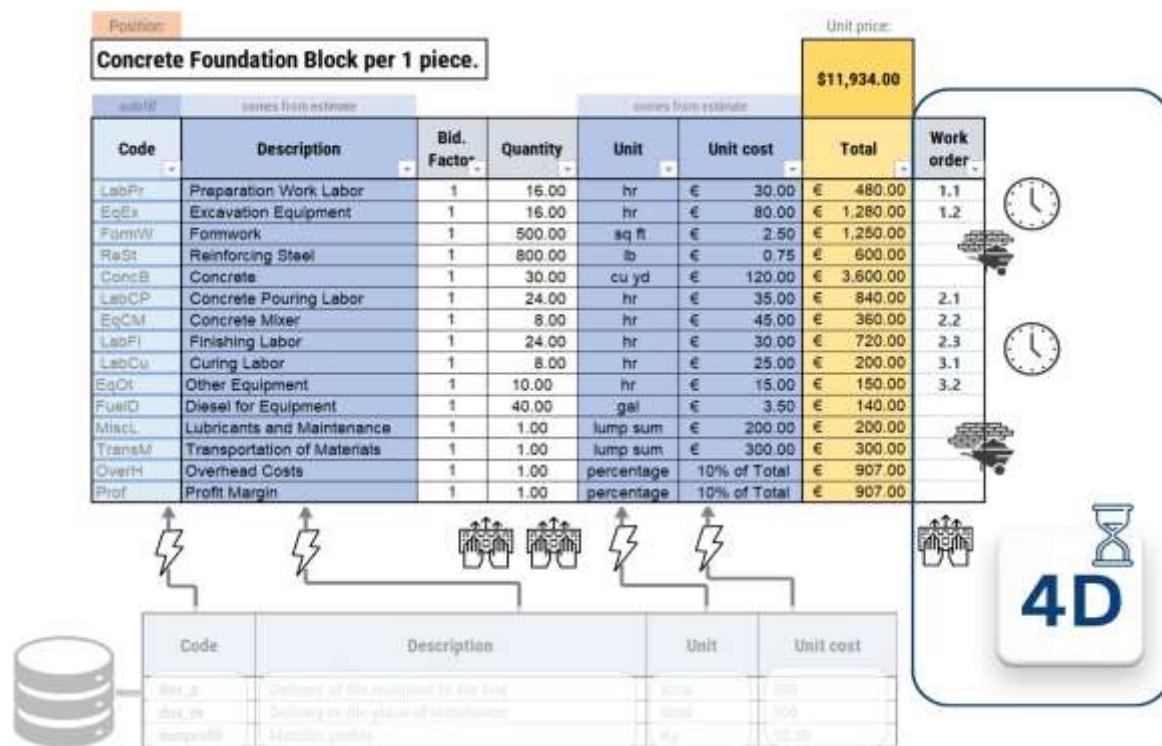


Fig. 5.3-1 Beregninger av arbeid i den ressursbaserte estimersmetoden inkluderer tidsfesting av arbeidstimer.

Beregninger på kostnadsnivå omfatter derfor ikke bare drivstoff- og materialkostnader (innkjøpskostnader), men også tiden som sjåfører, teknikere og hjelpearbeidere bruker på byggeplassen. I eksempelet (fig. 5.3-1) er kostnadstabellen en beregning av kostnadene ved å montere en grunnmur, inkludert de ulike

arbeidsfasene, som forarbeid, montering av ramme og støping av betong, samt materialer og arbeidskraft. Enkelte arbeidsoperasjoner, som for eksempel forberedelsesarbeid, kan imidlertid være uten materialekostnader, men kan inneholde betydelige midlertidige arbeidskostnader uttrykt i arbeidstimer.

For å planlegge arbeidssekvensen (for arbeidsplanen) på byggeplassen, legges attributtet "Arbeidsordre" manuelt til i kalkulasjonstabellen (fig. 5.3-1). Dette attributtet spesifiseres i en ekstra kolonne kun for elementer der måleenheten er uttrykt i tid (time, dag). Dette attributtet kommer i tillegg til arbeidskode, beskrivelse, mengde, måleenhet (parameter "Enhet") og kostnader. Den numeriske rekkefølgen (parameteren "Arbeidsrekkefølge") for aktiviteter gjør det mulig å angi i hvilken rekkefølge oppgavene skal utføres på byggeplassen og bruke den i planleggingen.

Byggeplan og automatisering av denne basert på kostnadsdata

Byggeplanen er en visuell fremstilling av arbeidsplanen og prosessene som skal utføres som en del av prosjektgjennomføringen. Den lages på grunnlag av detaljerte ressursberegninger (figur 5.3-1), der hver oppgave/jobb, i tillegg til ressurskostnadene, er planlagt etter tid og rekkefølge.

I motsetning til gjennomsnittsberegninger, der tidsestimater er basert på typiske timer for installasjon av materialer eller utstyr, er planleggingen i den ressursbaserte metoden basert på faktiske data i kostnadsberegningen. Hver post i estimatet knyttet til lønnskostnader er basert på den anvendte kalenderen, som tar hensyn til de faktiske forholdene for ressursbruk i løpet av arbeidsperioden. Justering av produktive timer gjennom koeffisienter på kalkulasjonsnivå (fig. 5.3-1 parameter "Bid. Faktor"), gjør det mulig å ta hensyn til forskjeller i produktivitet og sesongmessige særegenheter som påvirker tidspunktet for arbeidet.

For å finne start- og slutt datoene for byggeprosessen i Gantt-diagrammet tar vi verdiene for tidsattributtet for hvert element fra grunnmurskalkylen og multipliserer dem med antall blokker (i dette tilfellet antall grunnmurblokker av betong). Denne beregningen gir varigheten for hver oppgave. Deretter plotter vi disse varighetene inn på en tidslinje, med utgangspunkt i prosjektets startdato, for å lage en graf, og resultatet er en visuell fremstilling som viser når hver oppgave skal starte og avsluttes. Parameteren "Arbeidsrekkefølge" for prosesser gjør det i tillegg mulig for oss å forstå om arbeidsprosessen kjører parallelt ("Arbeidsrekkefølge", f.eks. 1.1-1.1) eller sekvensielt (1.1-1.2).

Gantt Chart er et grafisk verktøy for planlegging og prosjektstyring som representerer oppgaver som horisontale søyler på en tidslinje. Hver søyler representerer oppgavens varighet, start og slutt.

En tidsplan, eller et Gantt-diagram, hjelper prosjektledere og medarbeidere med å forstå når og i hvilken rekkefølge de ulike byggefaseene skal utføres, slik at ressursene utnyttes effektivt og tidsfristene overholdes.

La oss presentere en tidsplan for installasjonen av tre betongfundamentblokker ved hjelp av beregningene fra tabellen ovenfor. Bruk kostnadstabellen (fig. 5.3-1) fra eksempelet ovenfor og be LLM om å planlegge monteringen av de tre grunnmurblokrene, f.eks. til 1. mai 2024.

For å sende en kalkyle til LLM kan vi laste opp kalkyletabellen i XLSX-format eller ganske enkelt sette inn et skjermbilde av et JPEG-bilde av kalkylen direkte i LLM-chatten (fig. 5.3-1). LLM vil selv finne et bibliotek for å visualisere tabellbildet, og ved å multiplisere tidsattributtene for aktivitetene fra tabellen med volumet, vil LLM legge alle dataene inn i en graf.

 Send en tekstforespørsel til LLM:

Lag en arbeidsplan i form av et Gantt-diagram ved hjelp av verdiene fra den tidsrelaterte tabellen (fig. 5.3-1 JPEG i vedlegget) for installasjon av 3 fundamentblokker. Arbeidet for hver blokk skal utføres sekvensielt. Starten på arbeidene skal settes fra 01/05/2024 ↴

 LLMs svar

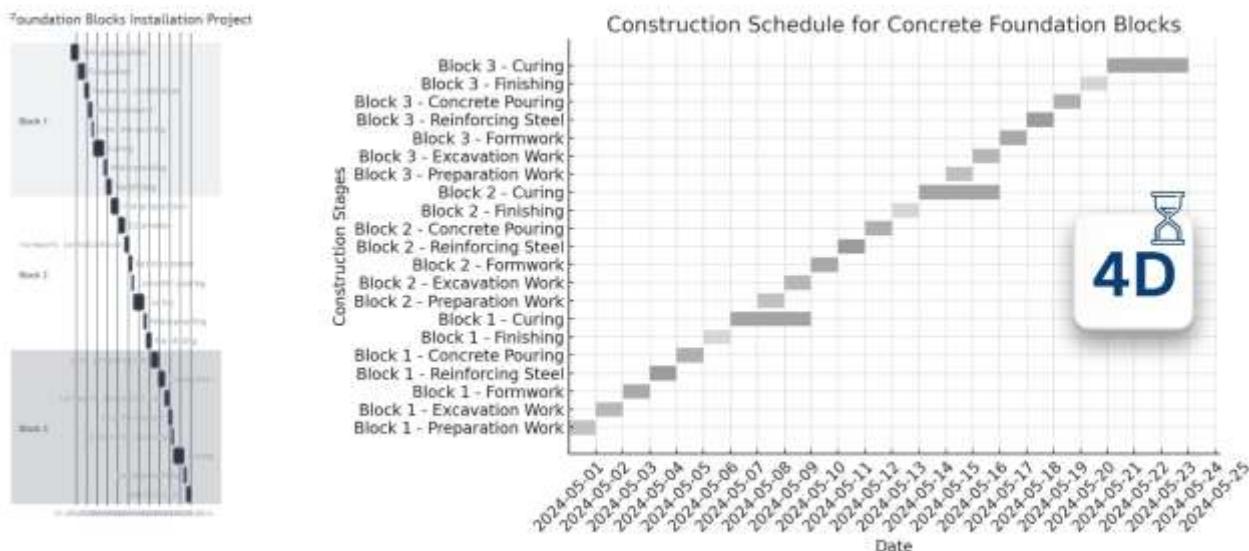


Fig. 5.3-2 Automatisk generert av flere LLM-er Gantt-diagrammet viser byggetrinnene for tre betongblokker, i henhold til betingelsene fra ledeteksten.

Den resulterende grafen (fig. 5.3-2) er et tidsdiagram der hver horisontale søyle tilsvarer et bestemt trinn i arbeidet med fundamentblokken og viser rekkefølgen av operasjoner (parameter "Arbeidsrekkefølge"), for eksempel forberedelse, utgraving, forskalingsinstallasjon, armering, betongstøping og etterbehandling,, dvs. de prosessene som har fylt ut tidsparametere og rekkefølge i beregningene.

En slik tidsplan (fig. 5.3-2) tar ikke hensyn til begrensninger knyttet til arbeidsdager, skift eller arbeidstidsstandarder, men er kun ment for konseptuell visualisering av prosessen. En nøyaktig tidsplan som gjenspeiler parallelliteten i arbeidet, kan suppleres med passende promter eller tillegginstruksjoner i chatterommet

Ved hjelp av en enkelt kalkyle (fig. 5.3-1), takket være volumattributter fra 3D -geometri, er det mulig å automatisk estimere både prosjektkostnaden via automatiserte -estimatorer og samtidig beregne tidskarakteristikken til grupper i form av tabeller eller grafer for ulike prosjektvarianter (fig. 5.3-3).

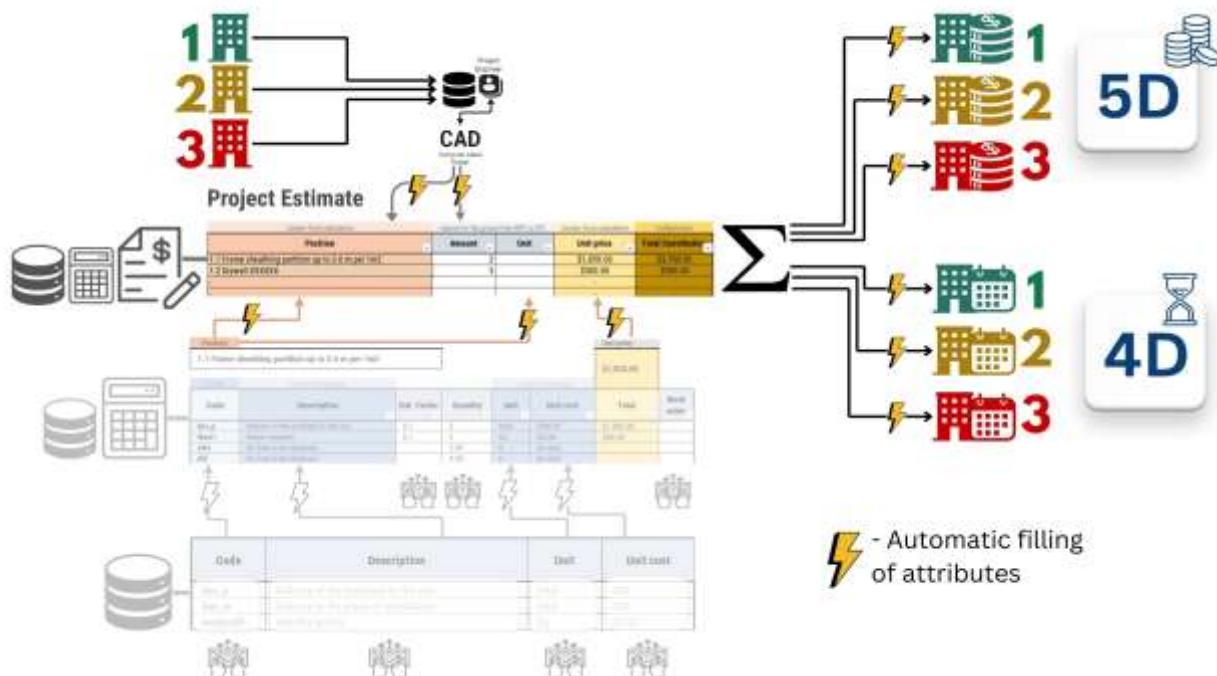


Fig. 5.3-3 Automatisk beregning, gjør det mulig å lage umiddelbare og automatiske kostnads- og tidsprognosenter ulike prosjektaalternativer.

Moderne modulbaserte ERP-systemer (fig. 5.4-4), som laster inn data fra CAD-modeller, bruker lignende automatiserte tidsberegningsmetoder som reduserer beslutningsprosessen betydelig. Dette gjør det mulig å planlegge arbeidsplaner umiddelbart og nøyaktig og beregne den totale tiden som kreves for å fullføre alle oppgavene i prosjektgjennomføringen, med hensyn til reelle priser.

Utvidede attributlag 6D -8D: fra energieffektivitet til sikkerhetssikring

6D, 7D og 8D er utvidede nivåer av informasjonsmodellering, som hver for seg bidrar med ytterligere lag av attributter til den omfattende prosjektinformasjonsmodellen, der grunnlaget er attributtene i 3D-modellen med deres antall og omfang. Hvert ekstra lag introduserer spesifikke parametere som er nødvendige for videre gruppering eller identifikasjon i andre systemer, som for eksempel eiendomsforvaltingssystemer (PMS), datastøttet anleggsstyring (CAFM), styring av byggeprosjekter (CPM) og sikkerhetssikringssystemer (SMS).



Figur 5.3-4 Attributtene 6D, 7D og 8D i datainformasjonsmodellen utvider hensynet til ulike aspekter ved prosjektet, fra energieffektivitet til sikkerhet.

- **6D** legges det til informasjon (attributtkolonner) om miljømessig bærekraft i tillegg til prosjektdatabasen (eller datarammen (figur 4.1-13)) med geometriske og volumetriske attributter for elementer. Dette omfatter informasjon om energieffektivitet, karbonfotavtrykk, materialers resirkulerbarhet og bruk av miljøvennlig teknologi. Disse dataene gjør det mulig å vurdere prosjektets miljøpåvirkning, optimalisere prosjektbeslutninger og oppnå målene for bærekraftig utvikling (ESG).
- **7D** kompletterer attributtene som kreves for å administrere vedlikehold av bygninger. Det dreier seg om data om vedlikeholdsplaner, komponenters livssyklus, teknisk dokumentasjon og reparasjonshistorikk. Dette settet med informasjon sikrer at modellen kan integreres med vedlikeholdssystemer (CAFM, AMS), muliggjør effektiv planlegging av vedlikehold, utskifting av utstyr og gir støtte gjennom hele anleggets livssyklus.
- **8D** ekstra attributtag, - inneholder informasjon knyttet til sikkerhet - både i byggefases og under senere drift. Modellen omfatter sikkerhetstiltak for personell, nødinstruksjoner, evakueringssystemer og krav til brannvern. Ved å integrere disse dataene i den digitale modellen kan man ta hensyn til risiko på forhånd og utvikle arkitektoniske, tekniske og organisatoriske løsninger som tar hensyn til helse- og sikkerhetskrav.

I strukturert tabellform representerer lagene 4D til 8D tilleggsattributter i form av kolonner med utfylte verdier (figur 5.3-5) som legges til de allerede utfylte 3D -modellattributtene som navn, kategori, type og volumetriske egenskaper. Verdiene i attributtagene 6D, 7D og 8D inneholder ytterligere tekstlige og numeriske data som resirkuleringsprosent, karbonfotavtrykk, garantiperiode, utskiftingssyklus, installasjonsdato, sikkerhetsprotokoller osv.



Figur 5.3-5 6D -8D legger til attributlag i datainformasjonsmodellen, som allerede inneholder geometriske og volumetriske attributter fra 3D -modellen.

For vårt nye vindu (figur 4.4-1) kan elementet med identifikatoren W-NEW (figur 5.3-5) ha følgende 3D -8D attributter:

3D -attributter - geometrisk informasjon hentet fra CAD -systemer:

- "Typenavn" - element "Vindu"
- "Bredde" - 120 cm
- I tillegg kan du legge til "Bounding Box"-punktene til et element eller dets "geometri BREP / MESH" som et eget attributt

Egenskapene til 6D - miljømessig bærekraft:

- Gjenvinningsgrad på 90 prosent
- "Karbonfotavtrykk - 1622 kg CO₂

Attributter 7D - data for objektstyring:

- "Garantiperiode" - 8 år
- "Replacement Cycle" er 20 år gammel
- "Vedlikehold" - kreves hvert år

Egenskapene til 8D - sørger for sikker bruk og drift av bygninger:

- "Installert"-vinduet - av selskapet "XYZ Windows"
- "Sikkerhetsstandard" - i samsvar med ISO 45001

Alle parametere som er registrert i en database eller et datasett (fig. 5.3-5), trengs av spesialister i ulike avdelinger for gruppering, søk eller beregninger. En slik flerdimensjonal attributbasert beskrivelse av prosjektobjekter gir et komplett bilde av livssyklusen, driftskravene og mange andre aspekter som er nødvendige for prosjektering, bygging og drift.

Estimering av CO₂ og beregning av karbondioksidutslipp fra byggeprosjekter

Ved siden av temaet bærekraft i byggeprosjekter på trinn 6D (figur 5.3-5) fokuserer moderne bygg- og anleggsvirksomhet på miljømessig bærekraft i prosjekter, der et av de viktigste aspektene blir vurdering og minimering av CO₂-utslipp som oppstår i løpet av prosjektets livssyklusfaser (f.eks. produksjon og installasjon).

Estimering og beregning av karbonutslipp fra byggematerialer er en prosess der det totale karbonutslippet bestemmes ved å multiplisere de volumetriske egenskapene til et element eller en gruppe elementer som brukes i et prosjekt, med en passende karbonutslippsfaktor for kategorien.

Å ta hensyn til karbonutslipp i vurderingen av byggeprosjekter som en del av de bredere ESG-kriteriene (miljø, sosiale forhold og selskapsstyring) gjør analysen enda mer kompleks. Dette er spesielt viktig for byggherrer og investorer som ønsker å oppnå relevante sertifiseringer som LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) eller DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Å oppnå en av disse sertifiseringene kan øke salgbarheten til en eiendom betraktelig, forenkle idriftsettelsen og sikre samsvar med kravene til bærekraftorienterte leietakere (ESG). Avhengig av prosjektkravene kan HQE (Haute Qualité Environnementale, den franske standarden for grønne bygninger), WELL (WELL Building Standard, med fokus på brukernes helse og komfort) GRESB (Global Real Estate Sustainability Benchmark) også brukes

ESGEnvironmental, Social and Governance) er et bredt sett med prinsipper som kan brukes til å vurdere en virksomhets selskapsstyring, sosiale og miljømessige påvirkning både internt og eksternt.

ESG, som opprinnelig ble utviklet på begynnelsen av 2000-tallet av finansfond for å gi investorer informasjon om brede miljømessige, sosiale og styringsmessige kriterier, har utviklet seg til å bli en viktig indikator for evaluering av både selskaper og prosjekter, inkludert byggeprosjekter. Ifølge undersøkelser utført av store konsulenterselskaper er hensynet til miljø, sosiale forhold og selskapsstyring (ESG) i ferd med å bli en integrert del av byggebransjen.

Ifølge EY (2023) "The Path to Carbon Neutrality" reduserer selskaper som aktivt implementerer ESG -prinsipper, ikke bare risikoen på lang sikt, men øker også effektiviteten i forretningsmodellene sine, noe som er spesielt viktig i den globale omstillingen av markedene [103]. PwCs ESG Awareness-rapport viser at bedrifters bevissthet om viktigheten av ESG - faktorer varierer fra 67% til 97%, og at de fleste organisasjoner ser på disse trendene som nøkkelen til fremtidig bærekraft [104], og at bedrifter for det meste opplever et betydelig press fra interessenter om å integrere ESG-prinsipper.

Integrering av ESG -prinsipper i byggeprosjekter bidrar dermed ikke bare til å oppnå internasjonale bærekraftsertifikater som LEED, BREEAM, DGNB, men sikrer også langsiktig bærekraft og konkurranseskyt for selskaper i bransjen.

En av de viktigste faktorene som påvirker det samlede karbonfotavtrykket til et byggeprosjekt, er

produksjons- og logistikkfasene for byggematerialer og komponenter. Materialene som brukes på byggeplassen, har ofte en avgjørende innvirkning på det totale CO₂-utsippet, spesielt i de tidlige fasene av prosjektets livssyklus - fra utvinning av råmaterialer til levering til byggeplassen.

Beregning av utsipp per kategori eller type bygningselement krever bruk av referansefaktorer for karbonutslipp som gjenspeiler mengden CO₂ som genereres fra produksjonen av ulike materialer. Slike materialer omfatter blant annet betong, murstein, resirkulert stål og aluminium. Disse verdiene hentes vanligvis fra anerkjente kilder og internasjonale databaser som UK ICE 2015 (Inventory of Carbon and Energy) og US EPA 2006 (U.S. Environmental Protection Agency) [105]. Tabellen nedenfor (figur 5.3-6) oppsummerer de grunnleggende utslippsfaktorene for en rekke vanlige byggematerialer. To nøkkelparametere er oppgitt for hvert materiale: spesifikke CO₂-utsipp (i kg per kg materiale) og omregningsfaktorer fra volum til vekt (i kg per kubikkmeter), som er nødvendige for å integrere beregningene i designmodellen og koble til QTO-datagrupperingen.



Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m ³ to kg
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

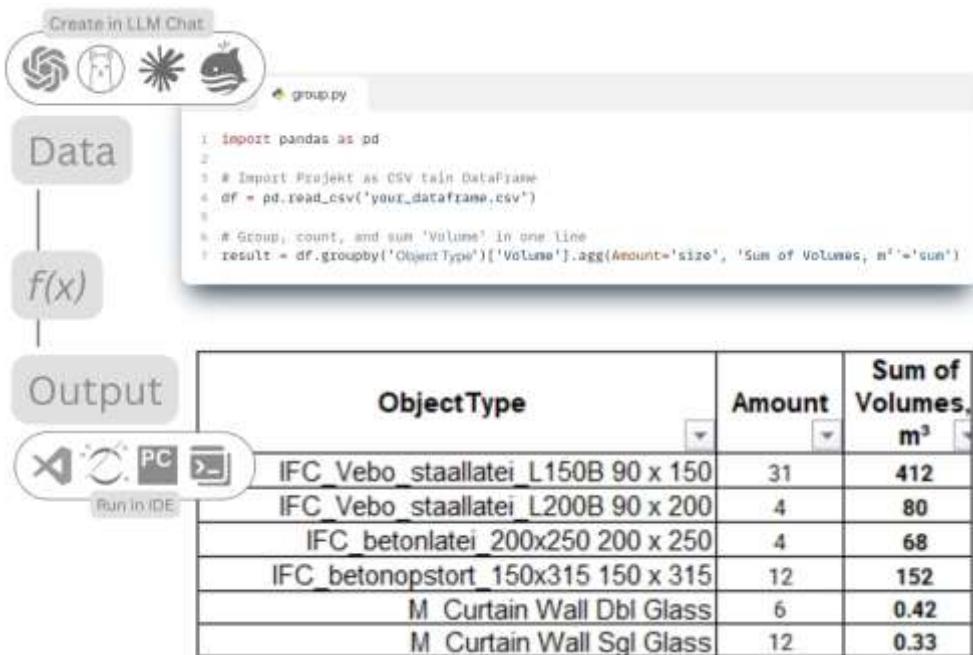
Figur 5.3-6 Mengde karbon som slippes ut under produksjonen av ulike byggematerialer, i henhold til databasen til UK ICE og US EPA.

For å beregne de totale CO₂-utsippene for et prosjekt, som med 4D og 5D -beregningene, er det nødvendig å bestemme attributtvolumene for hver gruppe av anlegg. Dette kan gjøres ved hjelp av kvantitative analyseverktøy (QTO) ved å hente ut attributtvolumene i kubikkmeter, som beskrevet i detalj i avsnittet om Quantity take-off. Disse volumene multiplisieres deretter med de aktuelle koeffisientene for attributtet "CO₂-prosessutsipp" for hver materialgruppe.

- La oss automatisk trekke ut volumtabellen etter elementtype fra CAD (BIM) -prosjektet ved å gruppere alle prosjektdataene slik vi allerede har gjort i de foregående kapitlene. For å utføre denne oppgaven, se LLM.

Grupper DataFrame-tabellen fra CAD (BIM) etter kolonneparametren "Object Name" (eller "Type") og vis antall elementer i hver gruppe, og oppsummer "Volume"-parametren for alle elementer i typen. ↴

💡 LLMs svar :

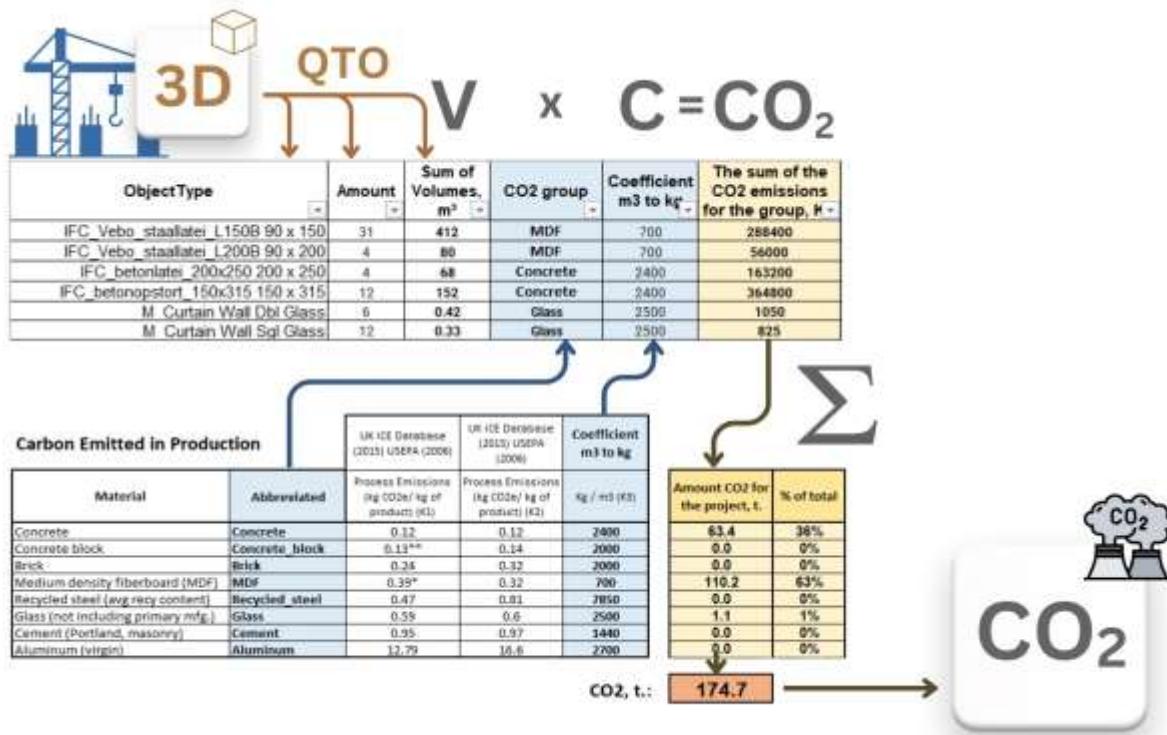


Figur 5.3-7 Den genererte koden i LLM har gruppert prosjektenhetene for oss etter type (ObjectType) med det oppsummerte attributtet "Volume".

For å automatisere beregningen av de totale CO₂-utslippe for hele prosjektet, er det bare å sette opp automatisk datatilordning i tabellen eller manuelt knytte elementtypene (figur 5.3-7) til de tilsvarende materialtypene (figur 5.3-6) fra utslippsfaktortabellen. Den ferdige tabellen med utslippsfaktorer og formler, samt kode for å hente volumer fra CAD-formater (BIM) og automatisere CO₂-bestemmelse, finner du på GitHub ved å søke etter "CO₂_calculating-the-embodyed-carbon. DataDrivenConstruction." [106].

Dataintegrering etter gruppering av QTO-elementer fra CAD-databasen muliggjør automatisk beregning av karbondioksidutslipp (fig. 5.3-8) for ulike designalternativer. Dette gjør det mulig å analysere virkningen av ulike materialer i forskjellige varianter og kun velge de løsningene som oppfyller krav til CO₂-utslipp for å oppnå et bestemt sertifikat når bygningen settes i drift

Estimering av CO₂-utslipp ved å multiplisere faktorer med volumet av grupperte prosjektelementer er et typisk eksempel på en oppgave i prosessen med å oppnå en ESG-klassifisering (f.eks. LEED-sertifisering) for et anlegg.

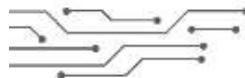


Figur 5.3-8 Integrering av QTO-grupper fra CAD-databaser gir nøyaktighet og automatisering ved beregning av endelige CO₂-utslipp.

På samme måte kan vi ved å definere elementgruppevolumer utføre beregninger for materialkontroll og logistikk, kvalitetsovervåking og -styring, energimodellering og -analyse og en rekke andre oppgaver for å oppnå en ny attributtstatus (parameter i tabellen) for både individuelle elementgrupper og hele prosjektet.

Hvis antallet slike beregningsprosesser i bedriften begynner å øke, oppstår spørsmålet om behovet for å automatisere slike beregninger og implementere beregningsresultatene i bedriftens prosesser og datahåndteringssystemer.

På grunn av kompleksiteten i en komplett løsning setter mellomstore og store selskaper i byggebransjen ut slik automatisering til systemutviklingsselskaper for ERP (eller PMIS). Utviklingsselskapene lager et enkelt, omfattende og modulbasert system for store kunder for å håndtere mange ulike informasjonslag, inkludert material- og ressursberegninger.



KAPITTEL 5.4.

CONSTRUCTION ERP OG PMIS SYSTEMER

Construction ERP -systemer på eksemplet med kalkyler og estimatorer

Modulære ERP-systemer integrerer ulike attributtag (informasjon) og datastrømmer i ett enkelt, omfattende system, slik at prosjektledere kan administrere ressurser, økonomi, logistikk og andre aspekter ved et prosjekt på en synkronisert måte innenfor én og samme plattform. Et ERP-system for bygg og anlegg fungerer som "hjernen" i byggeprosjekter, forenkler repeterende prosesser gjennom automatisering og sørger for åpenhet og kontroll gjennom hele byggeprosessen.

Construction ERP -systemer (Enterprise Resource Planning) er omfattende programvareløsninger som er utviklet for å styre og optimalisere ulike aspekter av byggeprosessen. Kjernen i ERP-systemer for bygg og anlegg er moduler for kostnadsberegning og planlegging, noe som gjør dem til et viktig verktøy for effektiv ressursplanlegging.

ERP-moduler -systemer gjør det mulig for brukerne å legge inn, behandle og analysere data på en strukturert måte som dekker ulike aspekter av et prosjekt, for eksempel material- og lønnskostnader, utstyrsutnyttelse, logistikkstyring, personalressurser, kontakter og andre byggeaktiviteter.

En av de funksjonelle blokkene i systemet er modulen for automatisering av forretningslogikk - BlackBox/WhiteBox, som spiller rollen som prosesskontrollsenter.

BlackBox /WhiteBox gjør det mulig for spesialister som bruker et ERP -system, å administrere ulike aspekter av virksomheten som allerede er forhåndskonfigurert av andre brukere eller administratorer, på en fleksibel måte ved hjelp av tilgangsrettigheter. I forbindelse med ERP-systemer refererer begrepene *BlackBox* og *WhiteBox* til graden av åpenhet og kontrollerbarhet i systemets interne logikk:

- **BlackBox** ("svart boks") - brukeren samhandler med systemet gjennom grensesnittet, uten tilgang til den interne logikken i prosessutførelsen. Systemet utfører beregninger på egen hånd, basert på forhåndsdefinerte regler som er skjult for sluttbrukeren. Han taster inn data og får resultatet uten å vite hvilke attributter eller koeffisienter som ble brukt inni.
- **WhiteBox** ("hvit boks") - prosesslogikken er tilgjengelig for visning, tilpasning og modifisering. Avanserte brukere, administratorer eller integratorer kan manuelt definere databehandlingsalgoritmer, beregningsregler og interaksjonsscenarioer mellom prosjektenheter.

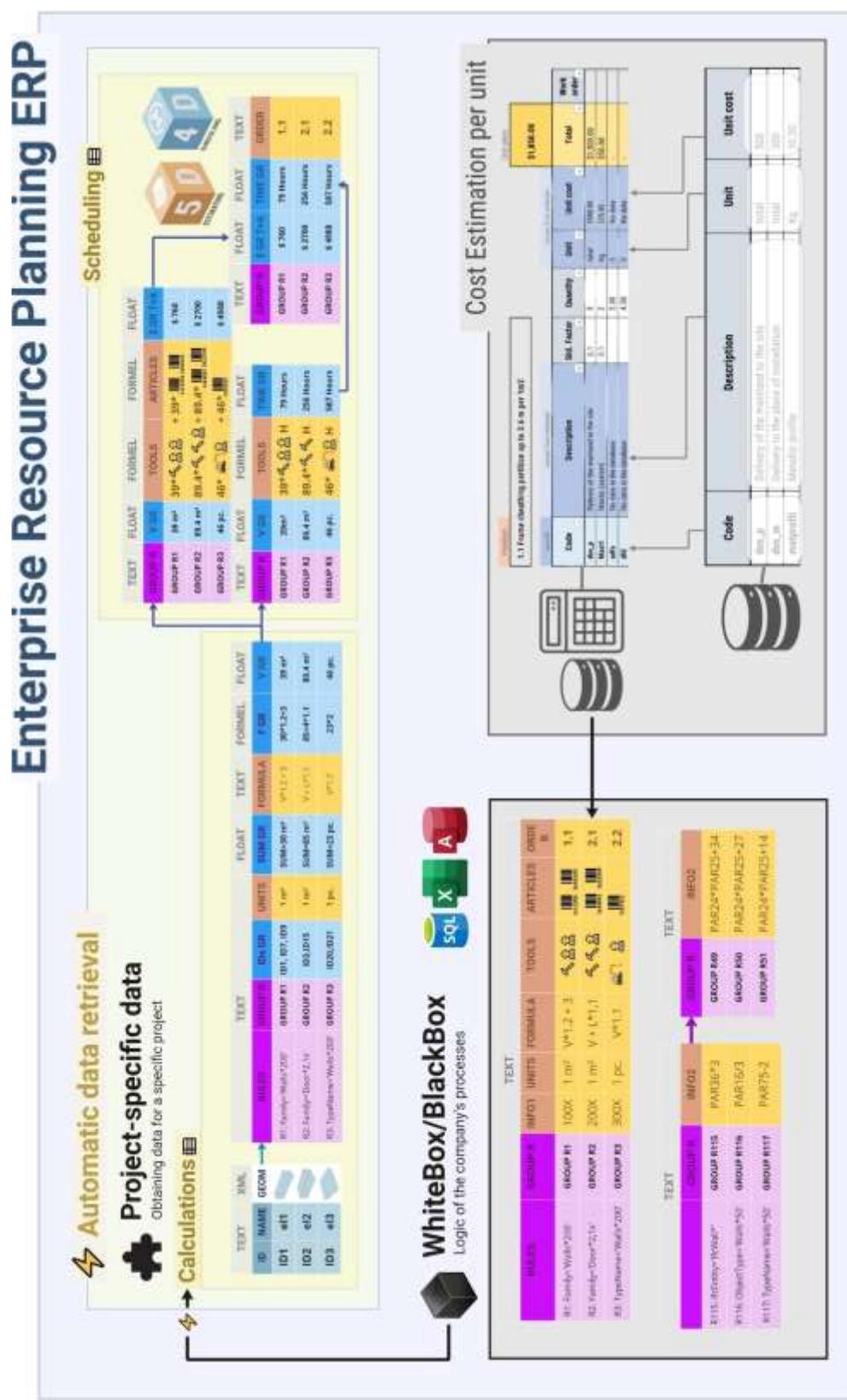


Fig. 5.4-1 Arkitektur for ERP -system for bygg- og anleggsbransjen, for å få estimerater og arbeidsplaner ved å fylle ut volumattributtene manuelt.

Et eksempel er når en erfaren bruker eller administrator setter opp en regel: hvilke attributter i et estimat som skal multipliseres med hverandre eller gruppert etter en bestemt egenskap, og hvor sluttresultatet skal registreres. Senere kan mindre erfarte fagfolk, for eksempel kalkulasjonsingeniører, ganske enkelt laste inn nye data i ERP-systemet via brukergrensesnittet - og få ferdige estimatorer, tidsplaner eller spesifikasjoner uten å måtte skrive kode eller forstå de tekniske detaljene i logikken.

I de foregående kapitlene ble beregnings- og logikkmodulene diskutert i forbindelse med LLM-interaksjon. I et ERP-miljø foregår slike beregninger og transformasjoner i moduler som er skjult bak et grensesnitt med knapper og skjemaer.

I det følgende eksempelet (fig. 5.4-1) har administratoren av ERP-systemet i BlackBox /WhiteBox-modulen definert regler for å matche attributter for enheter fra estimatorer med attributter for gruppert av QTO. Takket være denne BlackBox/WhiteBox-modulen som er konfigurert (av lederen eller administratoren), mottar brukeren (kalkulatøren eller ingeniøren), ved manuelt å legge til et mengde- eller volumattributt via ERP-brukergrensesnittet, automatisk de ferdige -kalkylene og arbeidsplanene. På denne måten blir beregnings- og estimatgenereringsprosessene som ble diskutert i de foregående kapitlene, ved hjelp av kode inne i ERP-systemet, til et halvautomatisk transportbånd.

Ved å koble denne halvautomatiserte prosessen til de volumetriske attributtene fra CAD (BIM)-modeller (fig. 4.1-13), for eksempel ved å laste CAD-prosjektet inn i ERP-modulen som er forhåndskonfigurert for dette formålet, blir dataflyten til en synkronisert mekanisme som er i stand til autonomt og øyeblikkelig å oppdatere verdien av individuelle grupper av elementer eller hele prosjektet som svar på eventuelle endringer i det i løpet av prosjekteringsfasen, når CAD-modellen lastes inn i ERP-systemet.

For å skape en automatisert dataflyt (fig. 5.4-2) mellom CAD (BIM) og ERP-systemer, må de grunnleggende prosessene og kravene for data fra CAD (BIM)-modelldatabaser defineres på en strukturert måte, slik vi allerede har diskutert i kapittelet "Krav og kvalitetssikring av data". Denne prosessen i ERP er delt inn i lignende trinn:

- **Opprettelse av valideringsregler (1)**, som spiller en viktig rolle når det gjelder å sikre at dataene som kommer inn i ERP-systemet, er korrekte. Valideringsreglene fungerer som filtre som validerer enheter og deres attributter, slik at bare de elementene som tilfredsstiller kravene, slipper inn i systemet. Du kan lese mer om verifisering og validering i kapittelet "Opprette krav og validere datakvalitet".
- Deretter følger en **verifiseringsprosess (2)** i ERP, som bekrefter at alle prosjektelementene med tilhørende attributter og verdier er opprettet på riktig måte og er klare for de neste behandlingstrinnene.
- Hvis det oppstår problemer med ufullstendige attributdata, **genereres det en rapport (3)**, og prosjektet sendes sammen med instruksjoner for korrigering til revisjon inntil det er klart for neste iterasjon.
- Når prosjektdataene er validert og verifisert, brukes de i en annen ERP-modul **(4) til å opprette Quantity Take-Off-tabeller (QTO)** som oppretter mengdeattributter for enhetsgrupper, materialer og ressurser i henhold til tidligere genererte regler (WhiteBox/BlackBox).
- Grupperte data etter matchingsregler eller QTO **integreses** automatisk **med beregninger (f.eks. kostnad og tid) (5)**.

- I det siste trinnet i ERP -systemet **genererer** brukeren **automatisk beregningsresultater (6)** (f.eks. kostnadsestimater, arbeidsplaner eller CO₂-utslipp) for hver enhetsgruppe og for prosjektet som helhet, ved å multiplisere scope-attributtene fra QTO-tabellen med attributtene i prosesstabellene (f.eks. estimerte elementer)

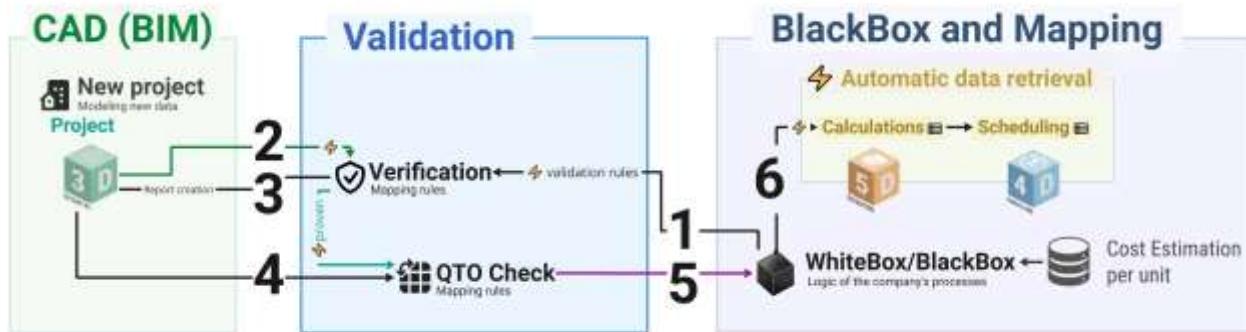


Fig. 5.4-2 Arkitektur for et ERP -system for bygg- og anleggsbransjen med CAD (BIM), fra opprettelse av valideringsregler (1) til automatisk beregning av kostnader og arbeidsplaner (5-6).

I et modulbasert ERP -system integreres prosessene ved hjelp av programvare som inkluderer et brukergrensesnitt. Bak grensesnittet ligger back-end, der strukturerte tabeller behandler data ved å utføre ulike operasjoner som er forhåndskonfigurert av lederen eller administratoren. Resultatet er at brukeren, takket være den forhåndsdefinerte og tilpassede automatiseringslogikken (i modulene BlackBox /WhiteBox), mottar halvautomatisk forberedte dokumenter som oppfyller oppgavene hans

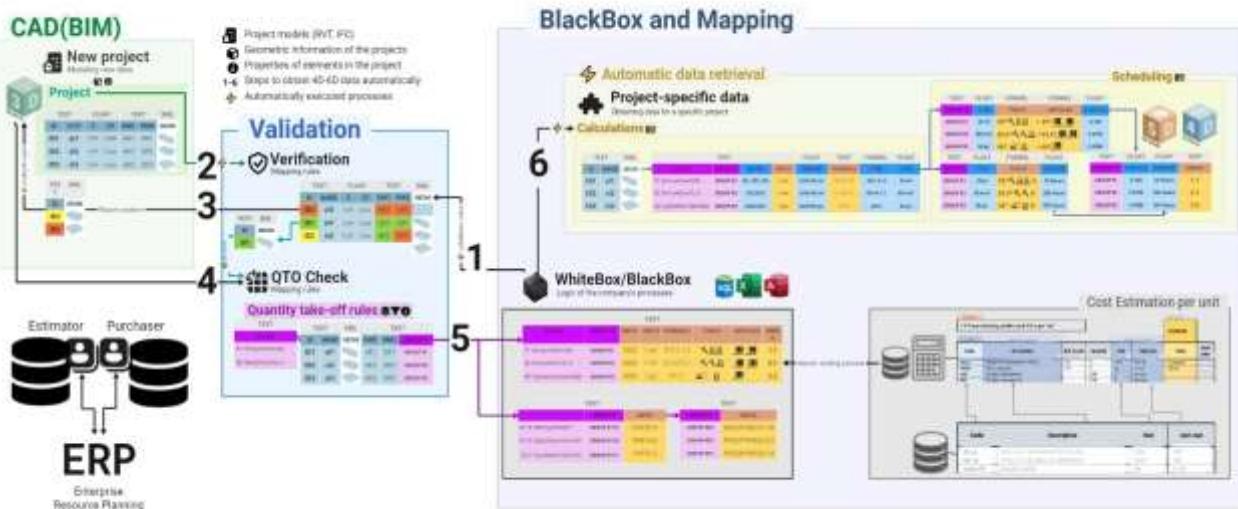


Fig. 5.4-3 ERP -systemet hjelper ledere og brukere med å flytte mellom spesialiserte tabeller for å generere nye data.

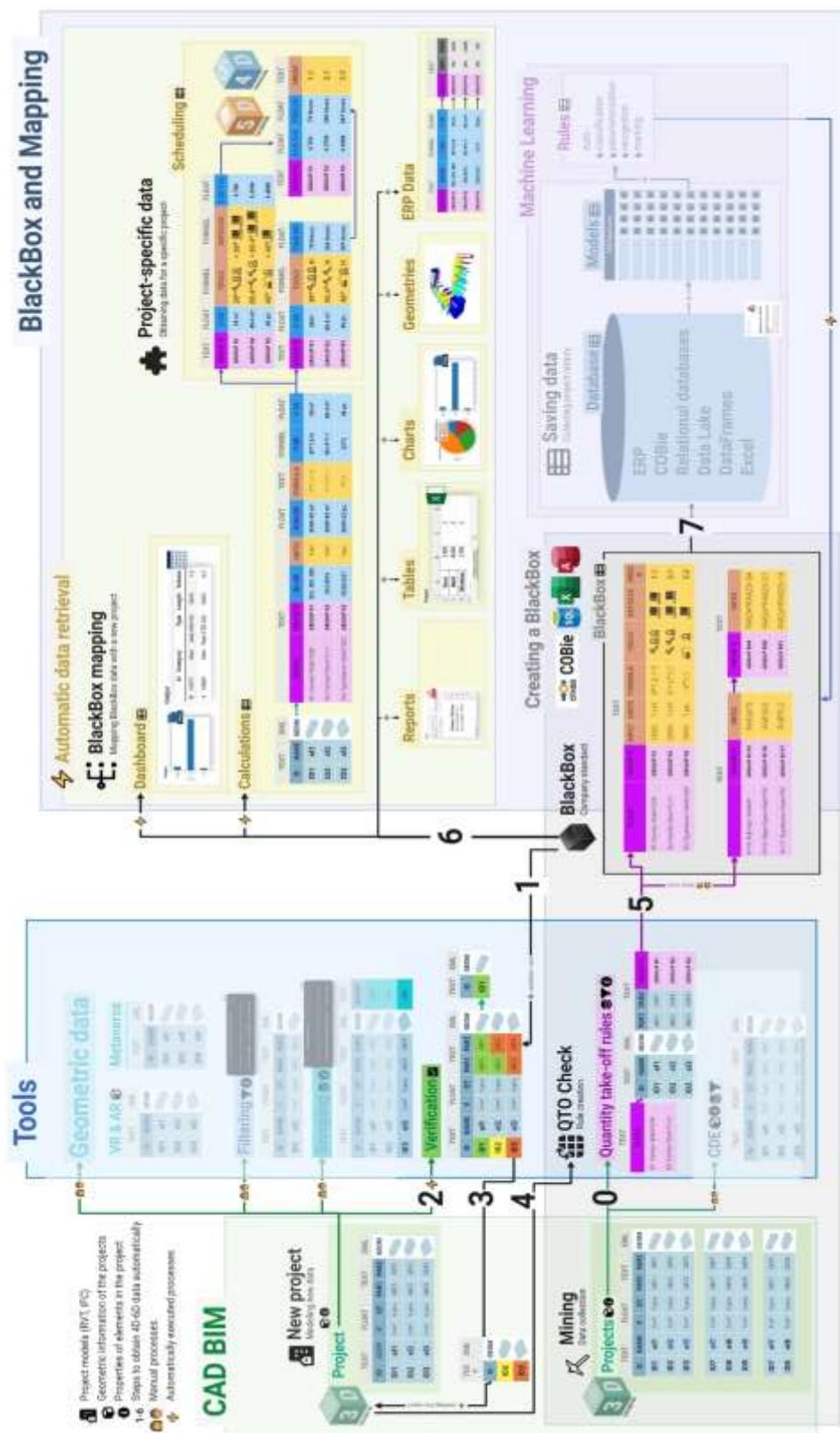


Fig. 5.4-4 ERP -systemet er integrert med analyseverktøy og automatiserer bedriftens beslutningsprosess.

På samme måte er prosessene i ERP -systemene, fra begynnelse til sluttberegning (trinn 1-6 Figur 5.4-3), en kjede av sammenhengende trinn som til syvende og sist gir åpenhet, effektivitet og nøyaktighet i planleggingen.

Moderne ERP-systemer for bygg- og anleggsbransjen inneholder ikke bare moduler for beregning av kostnader og tidsplaner, men også dusinvise av andre forhåndskonfigurerede moduler, som vanligvis dekker dokumenthåndtering, sporing av prosjektets fremdrift, kontraktstyring, forsyningsskjede og logistikk, samt integrering med andre forretningssystemer og plattformer. De integrerte analyseverktøyene i ERP gjør det mulig for brukerne å automatisere opprettelsen av dashbord for å overvåke prosjektets KPI-er (KPI - key performance indicators). Dette gir en sentralisert og konsekvent styring av alle aspekter ved et byggeprosjekt, og man forsøker å kombinere et stort antall applikasjoner og systemer på én enkelt plattform.

I fremtiden vil ERP -analytics bli brukt i kombinasjon med maskinlæring for å øke nøyaktigheten og optimalisere prosessen med å beregne fremtidige prosjekttattributter. Dataene og attributtene som analyseres og samles inn fra ERP-systemer i Big Data (fig. 5.4-4), vil i fremtiden danne grunnlaget for å lage prediktive modeller som nøyaktig kan forutse potensielle forsinkelser, risikoer eller for eksempel mulige endringer i materialekostnader.

Som et alternativ til ERP bruker byggebransjen ofte PMIS (Project Management Information System), et prosjektstyringssystem som er utviklet for detaljert kontroll av oppgaver på det enkelte byggeprosjekt.

PMIS: Mellomledd mellom ERP og byggeplassen

I motsetning til ERP, som dekker hele kjeden av en bedrifts forretningsprosesser, fokuserer PMIS på styring av et spesifikt prosjekt, overvåking av tidslinjer, budsjetter, ressurser og dokumentasjon.

PMIS (Project Management Information System) er en programvare for styring av byggeprosjekter som er utviklet for å planlegge, spore, analysere og rapportere om alle aspekter ved et prosjekt.

PMIS lar deg administrere dokumenter, tidsplaner, budsjetter og ved første øyekast kan PMIS se ut til å være en duplikatløsning av ERP, men den viktigste forskjellen er administrasjonsnivået:

- **ERP** fokuserer på forretningsprosessene i selskapet som helhet: kostnads-, kontrakts-, anskaffelses-, personal- og ressursstyring på konsernnivå.
- **PMIS** fokuserer på individuell prosjektleidelse, med detaljert planlegging, endringskontroll, rapportering og koordinering av deltakere.

I mange tilfeller er det ERP -systemene som allerede har tilstrekkelig funksjonalitet, og implementeringen av PMIS blir mer et spørsmål om bekvemmelighet og selskapets preferanser. Mange entreprenører og kunder bruker PMIS ikke fordi det er nødvendig, men fordi det er pålagt av leverandøren eller en stor kunde som ønsker å samle data på en bestemt plattform.

Det bør nevnes at det i den internasjonale terminologien for styring av byggeprosjekter finnes andre populære begreper som PLM (Product Lifecycle Management) og EPC og EPC-M (Engineering, Procurement and Construction Management) - kontraheringsmetoder i byggebransjen.

Hvis en bedrift allerede bruker ERP med prosjektstyringsmoduler, kan innføringen av PMIS være en

unødvendig kobling som dupliserer funksjonalitet. Men hvis prosessene ikke er automatiserte og dataene er fragmenterte, kan PMIS være et mer praktisk og vedlikeholdsvennlig verktøy.

Spekulasjon, profitt, isolasjon og mangel på åpenhet i ERP og PMIS

Til tross for enkle grensesnitt og prosedyrer er ERP- og PMIS-systemer for bygg- og anleggsbransjen i de fleste tilfeller lukkede og lite fleksible løsninger. Slike systemer leveres vanligvis som en forhåndskonfigurert programvarepakke fra én enkelt leverandør, med begrenset tilgang til interne databaser og prosesslogikk.

CAD-(BIM-) leverandører tar i økende grad over utviklingen og kontrollen av slike systemer, ettersom databasene deres inneholder den informasjonen som ERP-systemene trenger: kvantitative og volumetriske attributter for prosjektementer. Men i stedet for å gi tilgang til disse dataene i et åpent eller maskinlesbart format, tilbyr leverandørene kun begrensede brukerscenarier og lukket behandlingslogikk - forhåndsdefinert i BlackBox-moduler. Dette reduserer systemets fleksibilitet og gjør det umulig å tilpasse det til spesifikke prosjektforhold.

Begrenset datatransparens er fortsatt en av hovedutfordringene med digitale prosesser i byggebransjen. Lukket databasearkitektur, manglende tilgang til komplette sett med attributter for bygningselementer, fokus på BlackBox-automatiseringsmoduler og mangel på åpne grensesnitt øker risikoen for dokumentbyråkrati betydelig. Slike begrensninger skaper flaskehals i beslutningsprosessen, gjør det vanskelig å verifisere informasjon og åpner for at data skjules eller spekuleres i ERP/PMIS-systemene. Brukerne får vanligvis bare begrenset tilgang - enten det er et nedstrippet grensesnitt eller et delvis API - uten mulighet til å samhandle direkte med de primære datakildene. Dette er spesielt kritisk når det gjelder parametere som genereres automatisk fra CAD-prosjekter, for eksempel volumer, arealer og mengder som brukes til QTO-beregninger.

I stedet for å søke effektivitet gjennom prosessautomatisering, åpne data, redusere transaksjonskostnadene og skape nye forretningsmodeller, fokuserer mange byggefirmaer derfor på å styre eksterne parametere - manipulere faktorer, justeringsfaktorer og beregningsmetoder som påvirker prosjektkostnadene i lukkede ERP/PMIS-plattformer. Dette skaper rom for spekulasjon, forvrenger de reelle produksjonskostnadene og reduserer tilliten mellom alle deltakerne i byggeprosessen

I bygg- og anleggsbransjen utgjør fortjenesten forskjellen mellom inntektene fra et ferdigstilt prosjekt og de variable kostnadene, som omfatter design, materialer, arbeidskraft og andre direkte kostnader som er direkte knyttet til gjennomføringen av prosjektet. Det er imidlertid ikke bare teknologi eller logistikk som påvirker verdien av disse kostnadene, men også hvor raske og nøyaktige beregningene er, og kvaliteten på ledelsesbeslutningene i selskapet.

Problemet forverres av det faktum at kostnadsberegningsprosessene i de fleste byggefirmaer er ujennomsiktige, ikke bare for kundene, men også for de ansatte selv, som ikke er en del av kalkulasjons- eller økonomiavdelingene. En slik lukkethet bidrar til at det dannes en privilegert gruppe spesialister i

selskapet - bærere av "økonomisk ekspertise", som har enerett til å redigere attributter og korreksjonsfaktorer i ERP/PMIS-systemer. Disse medarbeiderne kan, sammen med bedriftslederne, faktisk kontrollere den økonomiske logikken i prosjektet.

Under slike forhold blir kalkulatørene til "økonomiske sjonglører", som må balansere mellom å maksimere selskapets fortjeneste og behovet for å opprettholde en konkurransedyktig pris for kunden. Samtidig må de unngå åpenbare og grove manipulasjoner for ikke å undergrave selskapets omdømme. Det er på dette stadiet koeffisienter fastsettes for å skjule overvurderte volumer eller kostnader for materialer og arbeider.

Som et resultat er hovedordningen for å øke effektiviteten og lønnsomheten til selskaper som opererer i byggebransjen ikke automatisering og akselerasjon av beslutningsprosesser, men spekulasjon i prisene på materialer og arbeider (fig. 5.4-5). Overvurdering av kostnadene for arbeider og materialer utføres av "grå" regnskap i lukkede ERP /PMIS - systemer ved å blåse opp prosentene over de gjennomsnittlige markedsprisene for materialer eller volum av arbeider ved hjelp av koeffisienter (fig. 5.1-6), som ble diskutert i kapittelet "Sammenstilling av beregninger og beregning av kostnadene for arbeider på grunnlag av ressursgrunnlaget".

Resultatet er at kunden får en kalkyle som ikke gjenspeiler den reelle kostnaden eller arbeidsomfanget, men som er et derivat av mange skjulte interne koeffisienter. Samtidig blir underleverandørene ofte tvunget til å kjøpe billigere materialer av lav kvalitet i et forsøk på å oppfylle de undervurderte prisene som hovedentreprenøren har satt, noe som forringer den endelige kvaliteten på byggingen.

spekulative prosessen med å søke profitt ut av løse luften ender opp med å skade både kundene som mottar upålitelige data, og lederne som tvinges til å finne stadig flere spekulasjonsmodeller.

Jo større prosjektet er, desto mer byråkratisk blir data- og prosesshåndteringen. Hvert trinn og hver modul skjuler ofte uggjennomsiktige koeffisienter og tillegg som er innebygd i beregningsalgoritmer og interne prosedyrer. Dette gjør det ikke bare vanskelig å revidere, men forvernger også det økonomiske bildet av prosjektet i betydelig grad. I store byggeprosjekter fører slik praksis ofte til en mangedobling (noen ganger opptil tidobling) av de endelige kostnadene, mens de reelle volumene og kostnadene forblir utenfor byggherrens effektive kontroll (fig. 2.1-3 Sammenligning av planlagte og faktiske kostnader i store infrastrukturprosjekter i Tyskland).

Ifølge McKinsey & Companys rapport *Imagining the Digital Future of Construction* (2016) blir store byggeprosjekter i gjennomsnitt fullført 20% senere enn planlagt og opptil 80% over budsjett [107].

Estimerings- og budsjetteringsavdelingene er i ferd med å bli det mest beskyttede ledet i et selskap. Tilgangen til dem er strengt begrenset, selv for interne spesialister, og på grunn av den lukkede logikken og databasestrukturene er det umulig å objektivt vurdere effektiviteten av prosjektbeslutninger uten forvrengninger. Mangelen på åpenhet fører til at bedriftene ikke blir tvunget til å optimalisere prosessene, men til å kjempe for å overleve ved hjelp av "kreativ" styring av tall og faktorer (fig. 5.3-1, fig. 5.1-6 - for eksempel parameteren "Bid. factor").



Fig. 5.4-5 Spekulasjonsrater på oppgjørsnivå er bedriftens viktigste fortjeneste og kunsten å sjonglere mellom kvalitet på arbeidet og omdømme.

Alt dette sår tvil om den fortsatte bruken av lukkede ERP/PMIS -systemer i byggebransjen. I sammenheng med den digitale transformasjonen og økende krav til åpenhet fra kundenes side (fig. 10.2-3), er det lite sannsynlig at prosjektgjennomføringen på lang sikt vil forblí avhengig av proprietære løsninger som begrenser fleksibiliteten, hindrer integrasjon og hemmer forretningsutviklingen.

Og uansett hvor gunstig det måtte være for byggefirmaer å jobbe med datasiloer og ugjennomsiktige data i lukkede databaser - fremtiden for byggebransjen vil unngåelig innebære en overgang til åpne plattformer, maskinlesbare og transparente datastrukturer og tillitsbasert automatisering. Denne omstillingen vil bli drevet frem fra toppen - under press fra kunder, tilsynsmyndigheter og samfunnet, som i økende grad krever ansvarlighet, bærekraft, åpenhet og økonomisk gjennomførbarhet.

Slutten på epoken med lukket ERP /PMIS: byggebransjen trenger nye tilnærminger

Bruken av store, modulbaserte ERP/PMIS -systemer som består av titalls millioner kodelinjer, gjør det ekstremt vanskelig å gjøre endringer i dem. I dette tilfellet blir overgangen til en ny plattform med moduler som allerede er forhåndskonfigurert for selskapet, titusenvis av artikler i ressursdatabaser (fig. 5.1-3) og tusenvis av ferdige kalkyler (fig. 5.1-6) en kostbar og langvarig prosess. Jo mer kode og eldre arkitektur - jo høyere blir den interne ineffektiviteten, og hvert nye prosjekt vil bare gjøre vondt verre. I mange bedrifter blir datamigrering og integrasjon av nye løsninger et flerårig epos med konstant omarbeiding og endeløse forsøk på å finne kompromisser. Resultatet er ofte at man vender tilbake til gamle, velkjente plattformer, til tross for deres begrensninger.

Som det fremheves i den tyske Black Book-rapporten [108] om systemsvikt i datahåndteringen i bygg- og anleggsbransjen, er fragmenteringen av informasjon og mangelen på en sentralisert tilnærming til håndteringen av den en viktig årsak til ineffektivitet. Uten standardisering og integrering mister data sin verdi og blir et arkiv i stedet for et styringsverktøy.

En viktig årsak til tap av datakvalitet er mangelfull planlegging og kontroll av byggeprosjekter, noe som ofte fører til betydelige kostnadsøkninger. I "Fokus: Kostnadseksplosjonen" analyserer Black Book de viktigste faktorene som bidrar til disse uønskede konsekvensene. Blant disse er mangelfulle behovsanalyser, manglende mulighetsstudier og ukoordinert planlegging som fører til merkostnader som kunne ha vært unngått.

I et modent IT-økosystem kan det å bytte ut et utdatert system sammenlignes med å bytte ut en bærende søyle i en allerede konstruert bygning. Det er ikke nok å bare fjerne den gamle og installere en ny - det er viktig å gjøre det på en slik måte at bygningen forblir stabil, takene ikke kollapser og all kommunikasjon fortsetter å fungere. Det er her vanskeligheten ligger: Enhver feil kan få alvorlige konsekvenser for hele bedriftssystemet.

Likevel fortsetter utviklere av store ERP-produkter for byggebransjen å bruke mengden skrevet kode som et argument for sin plattform. På fagkonferanser kan man fortsatt høre setninger som: "Det ville ta 150 årsverk å gjenskape et slikt system", til tross for at det meste av funksjonaliteten i slike systemer skjuler databaser og ganske enkle funksjoner for arbeid med tabeller, pakket inn i et spesielt fast brukergrensesnitt. I praksis blir kodevolumet på "150 årsverk" en byrde snarere enn et konkurransefortrinn. Jo mer kode - jo høyere supportkostnader, jo vanskeligere er det å tilpasse seg nye forhold, og jo høyere er inngangsteskelen for nye utviklere og kunder.

Mange modulbaserte byggesystemer ligner i dag tungvinte og utdaterte "Frankenstein-konstruksjoner" der enhver uforsiktig endring kan føre til feil. Hver ny modul øker kompleksiteten i et allerede overbelastet system og gjør det til en labyrinth som bare noen få spesialister kan forstå, noe som gjør det enda vanskeligere å vedlikeholde og modernisere.

Kompleksiteten blir også realisert av utviklerne selv, som med jevne mellomrom tar en pause for å refaktorisere - revidere arkitekturen for å ta hensyn til fremveksten av ny teknologi. Men selv om refaktorering gjøres regelmessig, vil kompleksiteten uunngåelig vokse. Arkitektene av slike systemer blir vant til den økende kompleksiteten, men for nye brukere og spesialister blir den en uoverstigelig barriere. Resultatet er at all ekspertisen konsentreres hos noen få utviklere, og systemet blir ikke lenger skalerbart. På kort sikt er slike eksperter nyttige, men på lang sikt blir de en del av problemet.

Organisasjoner vil fortsette å integrere "små" data med store data, og det er dumt å tro at én applikasjon - uansett hvor dyr eller robust den er - kan håndtere alt [109].

- Phil Simon, programleder for podcasten Conversations About Collaboration

Et legitimt spørsmål melder seg: Trenger vi virkelig slike tungvinte og lukkede systemer for beregning av kostnader og tidsbruk i form av tabeller, når andre bransjer for lengst har klart å løse tilsvarende oppgaver ved hjelp av analyseverktøy med åpne data og transparent logikk?

For øyeblikket er lukkede modulære plattformer fortsatt etterspurt i byggebransjen, først og fremst på grunn av de spesielle egenskapene ved kostnadsberegning (fig. 5.1-7). Slike systemer brukes ofte til å kjøre grå eller uujennomsiktige ordninger, slik at de reelle kostnadene kan skjules fra kunden. Etter hvert som

bransjen modnes digitalt, først og fremst kundene, og beveger seg inn i den såkalte "Uberised Era", vil imidlertid mellomleddene, det vil si byggefirmaene med sine ERP-systemer, miste sin betydning i tids- og kostnadsberegninger. Dette vil endre byggebransjens ansikt for alltid. Les mer i siste del av boken og i kapittelet "Construction 5.0: How to make money when you can't hide anymore".

Tusenvis av gamle løsninger som er akkumulert i løpet av de siste 30 årene, og som det er investert tusenvis av årsverk i å utvikle, vil begynne å forsvinne raskt. Overgangen til åpen, transparent og fleksibel datahåndtering er uunngåelig. Spørsmålet er bare hvilke selskaper som vil være i stand til å tilpasse seg disse endringene, og hvilke som vil forblive gisler i den gamle modellen.

En lignende situasjon er observert innen CAD-verktøy (BIM -), hvis data i dag fyller de volumetriske parametrene til prosjekteringsenheter i ERP/PMIS -systemer. Opprinnelig var ideen om BIM (utviklet allerede i 2002 [110]) basert på konseptet med én enkelt integrert database, men i praksis krever BIM-arbeid i dag et helt sett med spesialisert programvare og formater. Det som skulle forenkle design- og byggeledelse, har blitt til enda et lag med proprietære løsninger som kompliserer integrasjonen og reduserer fleksibiliteten i virksomheten.

Neste steg: effektiv bruk av prosjektdata

I denne delen har vi vist hvordan strukturerte data danner grunnlaget for nøyaktige kostnads- og tidsplanberegninger for byggeprosjekter. Automatisering av QTO-, planleggings- og kalkulasjonsprosessene reduserer arbeidskostnadene og forbedrer nøyaktigheten av resultatene betraktelig.

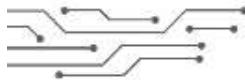
For å oppsummere denne delen er det verdt å fremheve de viktigste praktiske trinnene som vil hjelpe deg med å bruke de diskuterte tilnærmingene i dine daglige oppgaver. Disse tilnærmingene er universelle - de er nyttige både for den digitale transformasjonen av et selskap og for det daglige arbeidet til fagfolk som er involvert i beregninger:

- Automatiser rutineberegninger
 - Prøv å finne standard jobbkalkyler som du kan forholde deg til i arbeidet ditt
 - Analyser hvilke metoder som brukes til å kostnadsberegne eller kalkulere arbeider eller prosesser på byggeplassen i ditt land (figur 5.1-7).
 - Hvis du jobber med et CAD-system - lær deg funksjonene for automatisk uthenting av spesifikasjoner og QTO-data i CAD-programvaren din (BIM-).
 - Bruk LLM til å skrive utkast til kode for å automatisere beregninger
- Utvikle dine egne verktøy for QTO
 - Opprett skript eller tabeller for å automatisere volumtellingen
 - Standardisere kategorier og grupper av elementer for en konsekvent tilnærming til evaluering
 - Dokumentere beregningsmetodikken for å sikre reproducerte resultater i nye prosjekter
- Integriere ulike aspekter av prosjektet i arbeidet ditt

- Hvis du jobber med modulære systemer, bør du prøve å visualisere prosessene dine ikke bare i form av diagrammer eller diagrammer, men også på datanivå - spesielt i form av tabeller
- Behersk automatisk sammenslåing av data hentet fra CAD-databaser med beregninger - med Python-kode ved hjelp av gruppering, filtrering og aggregering
- Lag tydelige visualiseringer av QTO-grupper for å presentere kompleks informasjon for kolleger og kunder

Disse trinnene vil bidra til å bygge et bærekraftig kalkulasjonssystem basert på automatisering og standardisering av data. Denne tilnærmingen vil forbedre nøyaktigheten og redusere rutinen for daglige beregningsproblemer.

De følgende kapitlene fokuserer på de tekniske aspektene ved CAD - (BIM-) produkter og årsakene til at CAD-databaser fortsatt er vanskelige å integrere i bedriftenes forretningsprosesser. Hvis du ikke er interessert i historien om BIM-implementering i byggebransjen, utviklingen av DAK-verktøy og de tekniske aspektene ved å arbeide med disse teknologiene, kan du gå rett til den syvende delen av boken, "Datadrevet beslutningstaking".



MAKSIMAL BEKVEMMELIGHET MED DEN TRYKTE VERSJONEN

Du holder en gratis digital versjon av **Datadrevet konstruksjon**. For mer praktisk arbeid og rask tilgang til materialet anbefaler vi deg å ta hensyn til den **trykte utgaven**:



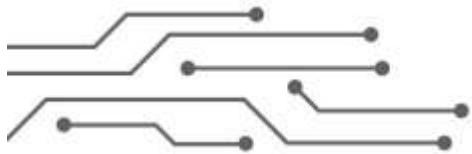
Alltid tilgjengelig: Boken i trykt format blir et pålitelig arbeidsverktøy, slik at du raskt kan finne og bruke de riktige visualiseringene og diagrammene i enhver arbeidssituasjon

Illustrasjoner av høy kvalitet: Alle bilder og all grafikk i den trykte utgaven presenteres i maksimal kvalitet

Rask tilgang til informasjon: enkel navigering, mulighet til å lage notater, bokmerker og arbeide med boken hvor som helst.

Ved å kjøpe den trykte versjonen av boken får du et praktisk verktøy for komfortabelt og effektivt arbeid med informasjon: muligheten til raskt å bruke visuelt materiale i daglige oppgaver, raskt finne de nødvendige ordningene og lage notater. I tillegg støtter kjøpet ditt formidling av åpen kunnskap.

Bestill en trykt versjon av boken på: datadrivenconstruction.io/books



VI DEL

CAD OG BIM: MARKEDSFØRING, VIRKELIGHET OG FREMTIDEN FOR DESIGNDATA I BYGG- OG ANLEGSBRANSJEN

Den sjette delen av boken presenterer en kritisk analyse av utviklingen av CAD og BIM -teknologiene og deres innvirkning på datahåndteringsprosessene i byggebransjen. Den spører den historiske utviklingen av BIM-konseptet fra den opprinnelige ideen om en integrert database til dagens markedsføringskonstruksjoner som fremmes av programvareleverandører. Vi ser nærmere på hvordan proprietære formater og lukkede systemer påvirker effektiviteten i håndteringen av prosjektdata og den generelle ytelsen i byggebransjen. Kompatibilitetsproblemene mellom ulike CAD-systemer og vanskelighetene med å integrere dem i byggfirmaenes forretningsprosesser blir analysert i detalj. Aktuelle trender i retning av forenklede åpne dataformater, som USD, og deres potensielle innvirkning på bransjen blir diskutert. Alternative metoder for å hente ut informasjon fra lukkede systemer presenteres, blant annet reverse engineering-teknikker. Utsiktene for bruk av kunstig intelligens og maskinlæring for å automatisere design- og dataanalyseprosesser i bygg- og anleggsbransjen blir analysert. Prognosør for utviklingen av designteknologier som er orientert mot brukernes reelle behov snarere enn programvareleverandørenes interesser, blir formulert.

KAPITTEL 6.1.

FREMVEKSTEN AV BIM- KONSEPTER I BYGGEBRANSJEN

Opprinnelig var denne sjette delen, som er viet CAD (BIM), ikke inkludert i den første versjonen av boken. Temaene proprietære formater, geometriske kjerner og lukkede systemer er altfor tekniske, overleddet med detaljer og tilsvarende ubrukelige for dem som bare ønsker å forstå hvordan man jobber med data. Tilbakemeldinger og forespørsler om å utdype den første versjonen av boken viste imidlertid at uten å forstå kompleksiteten i CAD-systemenes indre funksjoner, geometriske kjerner, de mange formatene og inkompatible lagringssystemene for de samme dataene, er det umulig å virkelig forstå hvorfor konseptene som leverandørene fremmer, ofte gjør det vanskelig å arbeide med informasjon og hindrer overgangen til åpen parametrisk design. Derfor har denne delen fått sin egen plass i bokens struktur. Hvis DAK (BIM) ikke er en prioritert for deg, kan du hoppe rett til neste del - "DEL VII: Datadrevet beslutningstaking, analyse, automatisering og maskinlæring".

Historien bak fremveksten av BIM og åpen BIM som markedsføringskonsepter for CAD- leverandører

Med inntoget av digitale data på 1990-tallet ble dataeknologi introdusert ikke bare i forretningsprosesser, men også i designprosesser, noe som førte til konsepter som CAD (computer-aided design systems) og senere BIM (building information modelling)

Men som alle andre innovasjoner er de ikke utviklingens sluttspunkt. Konsepter som BIM har blitt en viktig milepæl i byggebransjens historie, men før eller senere vil de måtte vike plassen for bedre verktøy og tilnærmingar som bedre vil kunne møte fremtidens utfordringer.

Overveldet av innflytelsen fra CAD-leverandørene og forvirret av kompleksiteten i sin egen implementering, kan det godt hende at konseptet BIM, som dukket opp i 2002, ikke overlever trettiårsjubileet, som en rockestjerne som blusset sterkt opp, men som raskt bleknet. Årsaken er enkel: kravene fra dataforskerne endrer seg raskere enn CAD-leverandørene kan tilpasse seg dem.

På grunn av mangel på kvalitetsdata krever dagens fagfolk i byggebransjen interoperabilitet på tvers av plattformer og tilgang til åpne data fra CAD- prosjekter for å forenkle analyse og behandling. Kompleksiteten i CAD-data og den forvirrende behandlingen av CAD-data har en negativ innvirkning på alle som er involvert i byggeprosessen: designere, prosjektledere, bygningsarbeidere på byggeplassen og, til syvende og sist, kunden.

I stedet for et fullverdig datasett for drift mottar kunden og investoren i dag containere i CAD- -formater som krever komplekse geometriske kjerner, forståelse av dataskjemaer, årlig oppdatert API -dokumentasjon og spesialisert CAD-programvare (BIM) for å arbeide med dataene. Samtidig forblir mye av designdataene ubrukt.

I dagens design- og konstruksjonsverden fører kompleksiteten i tilgangen til DAK-data til at prosjektstyringen blir overkonstruert. Mellomstore og store selskaper som arbeider med DAK-data eller

utvikler BIM -løsninger, er enten tvunget til å opprettholde tette relasjoner med DAK-leverandører -løsninger for å få tilgang til data via API-er, eller omgå DAK-leverandørens begrensninger ved å bruke dyre SDK-konvertere for å reversere, for å få tilgang til åpne data [75].

Den proprietære datatilnærmingen er utdatert og oppfyller ikke lenger kravene i dagens digitale miljø. Fremtiden vil dele selskapene i to typer: de som bruker åpne data effektivt, og de som forlater markedet.

Konseptet BIM (Building Information Modeling) dukket opp i byggebransjen da en av de største CAD-leverandørene publiserte Whitepaper BIM [54] i 2002, og ble supplert med det maskintekniske konseptet BOM (Bills of Materials), som har sitt utspring i den parametriske tilnærmingen til opprettelse og behandling av prosjektdata (fig. 6.1-1). Den parametriske tilnærmingen til opprettelse og behandling av konstruksjonsdata var en av de første som ble implementert i Pro-E-systemet for maskinkonstruksjon (MCAD). Dette systemet ble en prototyp [111] for mange moderne CAD-løsninger, inkludert de som brukes i byggebransjen i dag.

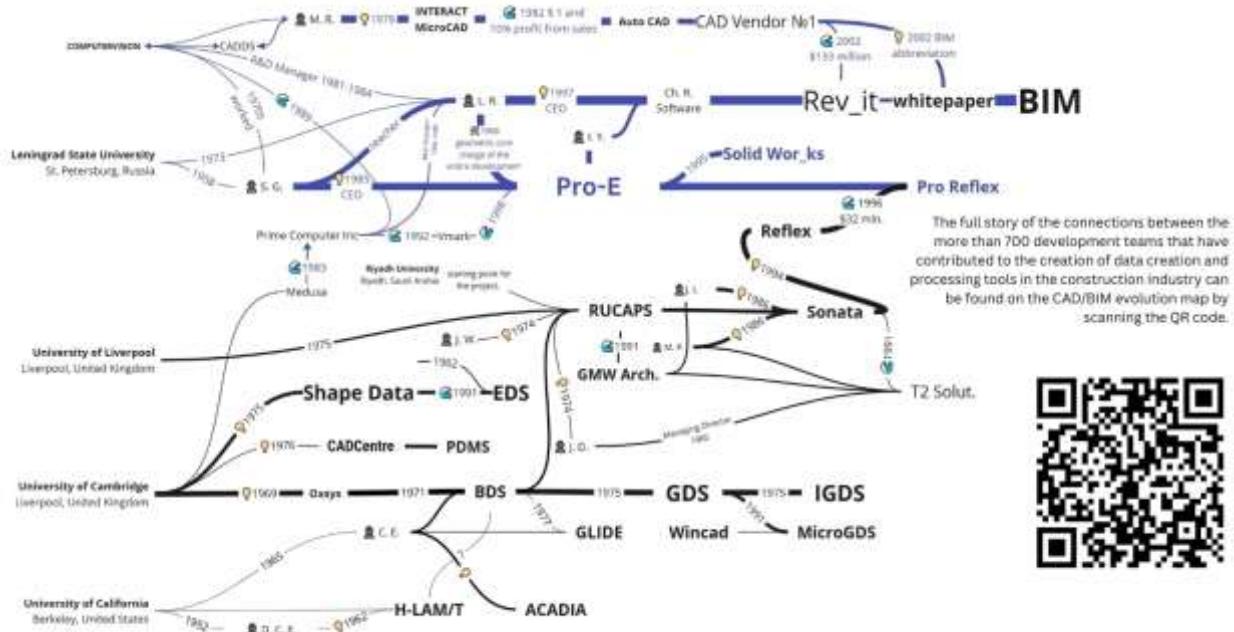


Fig. 6.1-1 Kart over historien til BIM-konseptet og lignende konsepter.

Journalister og AEC-konsulenter, som promoterte CAD-verktøy -leverandører frem til begynnelsen av 2000-tallet, rettet oppmerksomheten mot Whitepaper BIM fra 2002. Det var Whitepaper BIM 2002-2004 og artiklene som ble publisert i 2002, 2003, 2005 og 2007, som spilte en nøkkelrolle i populariseringen av BIM-konseptet i byggebransjen [112].

Bygningsinformasjonsmodellering er en strategi..... [navn på CAD-leverandør] for å anvende informasjonsteknologi i byggebransjen.

- BIM Whitepaper, 2002 [60].

På midten av 2000-tallet begynte "forskere" å knytte BIM- -konseptet som ble publisert av CAD- leverandøren i 2002, til tidligere vitenskapelige arbeider, som Charles Eastmans BDS, som ble grunnlaget for systemer som GLIDE, GBM, BPM, RUCAPS. I sitt banebrytende verk Building Description System (1974) la Charles Eastman det teoretiske grunnlaget for moderne informasjonsmodellering. Begrepet "database " dukker opp 43 ganger i hans arbeid (fig. 6.1-2) - oftere enn noe annet, bortsett fra ordet "building".

Eastmans hovedidé var at all informasjon om en bygning - fra geometri til elementenes egenskaper og deres innbyrdes forhold - skulle lagres i én strukturert database. Det er fra denne databasen at tegninger, spesifikasjoner, beregninger og overholdelse av regler kan genereres og analyseres automatisk. Eastman kritiserte eksplisitt tegninger som en utdatert og overflødig kommunikasjonsmetode, og pekte på duplisering av informasjon, problemer med oppdatering og behovet for manuelle oppdateringer når det gjøres endringer. I stedet foreslo han én enkelt digital modell i en database, der enhver endring gjøres én gang og automatisk gjenspeiles i alle visninger.

Det er verdt å merke seg at Eastman i sitt konsept ikke satte visualisering i sentrum. Informasjon var det sentrale i systemet hans: parametere, relasjoner, attributter, analyse- og automatiseringsmuligheter. Tegninger var bare én av flere måter å vise data fra databasen på, og ikke den primære kilden til designinformasjon.

I det første Whitepaper on BIM fra den ledende CAD-leverandøren ble uttrykket "database " brukt like ofte som i Charles Eastmans BDS - 23 ganger [60] over sju sider, og var et av de mest populære ordene i dokumentet etter "Building", "Information", "Modeling" og "Design". I 2003 dukket imidlertid begrepet "database" bare opp to ganger i lignende dokumenter [61], og på slutten av 2000-tallet var temaet databaser praktisk talt forsvunnet fra diskusjonen om designdata. Konseptet med "en enkelt integrert database for visuell og kvantitativ analyse" ble derfor aldri helt realisert.

Byggebransjen har altså gått fra Charles Eastmans progressive BDS-konsept med vekt på databaser og Samuel Geisbergs ideer om automatisk oppdatering av konstruksjonsdata fra databaser i det maskintekniske produktet Pro-E (forgjengeren til de populære CAD -løsningene som brukes i byggebransjen i dag) til dagens markedsførte BIM, der datahåndtering gjennom databaser knapt nevnes, til tross for at det var dette som var tanken bak det opprinnelige teoretiske konseptet

- **Data tilhører leverandørene, ikke brukerne.** Prosjektinformasjon er låst i proprietære formater eller skytjenester i stedet for å være tilgjengelig i åpne og uavhengige formater.

Konstruktører og prosjektledere har ofte verken tilgang til CAD-databasen -systems eller til formatet som deres egne prosjektdata er lagret i. Dette gjør det umulig å raskt verifisere informasjon eller formulere krav til datastruktur og -kvalitet (fig. 6.1-3). Tilgang til slike data krever et helt sett med spesialiserte programmer som er koblet sammen via API-er og plug-ins, noe som fører til en overdreven byråkratisering av prosessene i byggebransjen. Samtidig brukes disse dataene av dusinvise av informasjonssystemer og hundrevis av spesialister.

Vi må kunne håndtere alle disse dataene [CAD (BIM)], lagre dem digitalt og selge programvare for livssyklus- og prosessstyring, for for hver ingeniør [designer] som skaper noe [i CAD-programvare], er det ti personer som jobber med disse dataene" [41].

- Administrerende direktør i CAD - leverandøren som skapte BIM-konseptet, 2005.

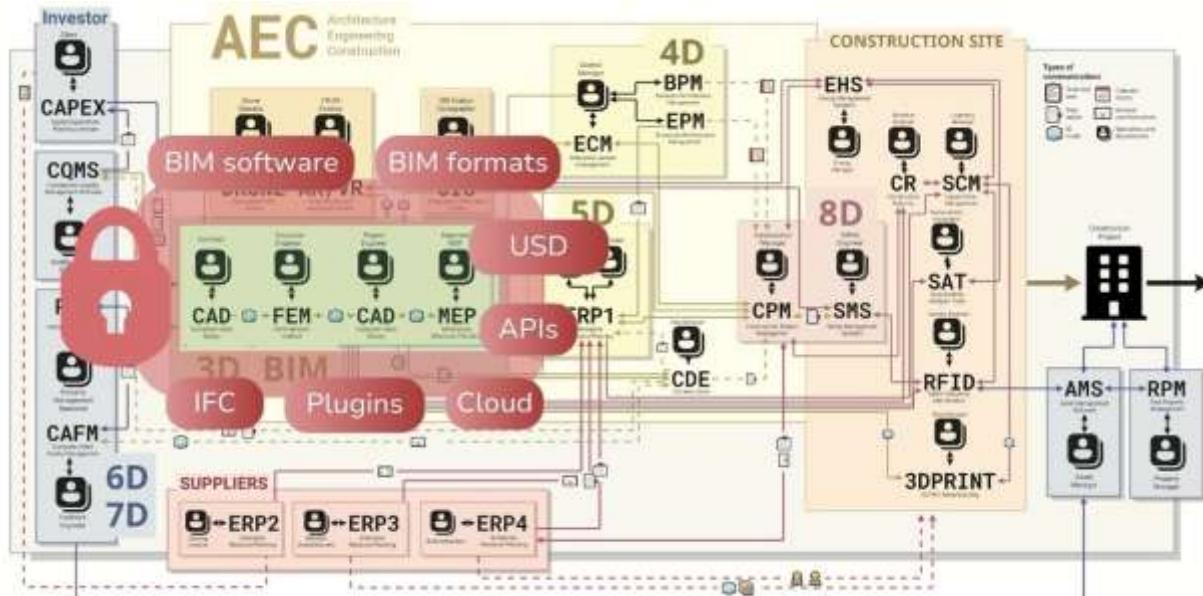


Fig. 6.1-3 CAD- (BIM-) databaser er fortsatt et av de siste lukkede systemene for IT-avdelinger og dataforvaltere i bygg- og anleggsbransjens økosystem.

Når det blir tydelig at BIM mer er et middel til å kommersialisere databaser enn et fullverdig databasehåndteringsverktøy, melder det seg et logisk spørsmål: Hvordan kan vi gjenvinne kontrollen over dataene? Svaret er å bruke åpne datastrukturer der brukeren, og ikke programvareleverandøren, blir eier av informasjonen.

Brukere og løsningsutviklere i byggebransjen vil, i likhet med sine kolleger i andre bransjer, unngåelig bevege seg bort fra den vage programvareleverandørterminologien som har dominert de siste 30 årene, og

i stedet fokusere på de viktigste aspektene ved digitalisering - "data" og "prosesser".

På slutten av 1980-tallet ble den digitale utviklingen i byggebransjen først og fremst fremstilt som et spørsmål om datatilgang og styring av prosjektinformasjon. Over tid har imidlertid fokuset endret seg. I stedet for å utvikle transparente og tilgjengelige metoder for å arbeide med data, ble IFC-formatet og det åpne BIM-konseptet aktivt promotert som forsøk på å avlede spesialistenes oppmerksomhet fra temaet prosjektdatabasehåndtering.

Fremveksten av det åpne formatet IFC i byggebransjen

Det såkalte åpne formatet IFC (Industry Foundation Classes) er posisjonert som en standard for å sikre interoperabilitet mellom ulike CAD-systemer (BIM -). Utviklingen av formatet ble utført innenfor rammen av organisasjoner som ble opprettet og kontrollert av store CAD-leverandører. Basert på IFC-formatet utviklet to CAD- selskaper i 2012 markedsføringskonseptet OPEN BIM[63].

IFC (Industry Foundation Classes) er en åpen standard for datautveksling i byggebransjen, utviklet for å sikre interoperabilitet mellom ulike CAD - (BIM-) systemer.

Open BIM - Konseptet innebærer at man arbeider med informasjon fra CAD-databaser og utveksler informasjon mellom systemer via et åpent format for utveksling av CAD-data - IFC.

Open BIM-programmet er en markedsføringskampanje initiert av... [1 DAK-leverandør],... [2 DAK-leverandør] og andre selskaper for å oppmuntre til og legge til rette for global koordinert markedsføring av OPEN BIM-konseptet i hele AEC-bransjen, med konsekvent kommunikasjon og felles merkevarebygging tilgjengelig for deltakerne i programmet.

- Fra CAD-leverandørens nettsted, OPEN BIM-programmet, 2012 [113].

IFC ble tilpasset av det tekniske universitetet i München fra det maskintekniske formatet STEP på slutten av 1980-tallet, og ble senere registrert av et stort designfirma og en stor DAK- leverandør for å danne IAI (Industry Alliance for Interoperability) i 1994 [114] (fig. 6.1-4). Formatet IFC ble utviklet for å sikre interoperabilitet mellom ulike DAK-systemer og var basert på prinsippene i det maskintekniske formatet STEP, som i sin tur sprang ut av formatet IGES som ble opprettet i 1979 av en gruppe DAK-brukere og -leverandører med støtte fra NIST (The National Institute of Standards and Technology) og det amerikanske forsvarsdepartementet [115].

Den komplekse strukturen til IFC, den nære avhengigheten av den geometriske kjernen, samt forskjeller i implementeringen av formatet i ulike programvareløsninger har imidlertid ført til mange problemer i den praktiske anvendelsen. Lignende problemer - tap av detaljer, begrenset nøyaktighet og behovet for å bruke mellomformater - ble tidligere møtt av spesialister innen maskinteknikk når de arbeidet med IGES- og STEP-formater, som IFC sprang ut av.



Fig. 6.1-4 Kart over koblinger mellom utviklingsteam og CAD-produkter (BIM) [116].

I 2000 publiserte den samme CAD-leverandøren som registrerte IFC-formatet og opprettet IAI-organisasjonen (senere bS), hvitboken "Integrated Design and Manufacturing: Benefits and Rationale" [65]. Her understrekkes viktigheten av å opprettholde full datagranularitet ved utveksling mellom programmer innenfor samme system, uten å bruke nøytrale formater som IGES, STEP [identisk med IFC]. I stedet ble det foreslått at applikasjonene skulle ha direkte tilgang til den underliggende CAD-databasen for å forhindre tap av nøyaktig informasjon.

I 2002 kjøper den samme CAD-leverandøren produktet parametrisk stykkliste (fig. 3.1-18, flere detaljer i tredje del) og danner på grunnlag av dette BIM-konseptet. Som et resultat av dette brukes bare lukkede CAD-formater eller IFC-formatet (STEP) i utvekslingen av byggeprosjektdata, hvis begrensninger ble skrevet om av CAD-leverandøren selv i 2000, som brakte dette formatet til byggebransjen.

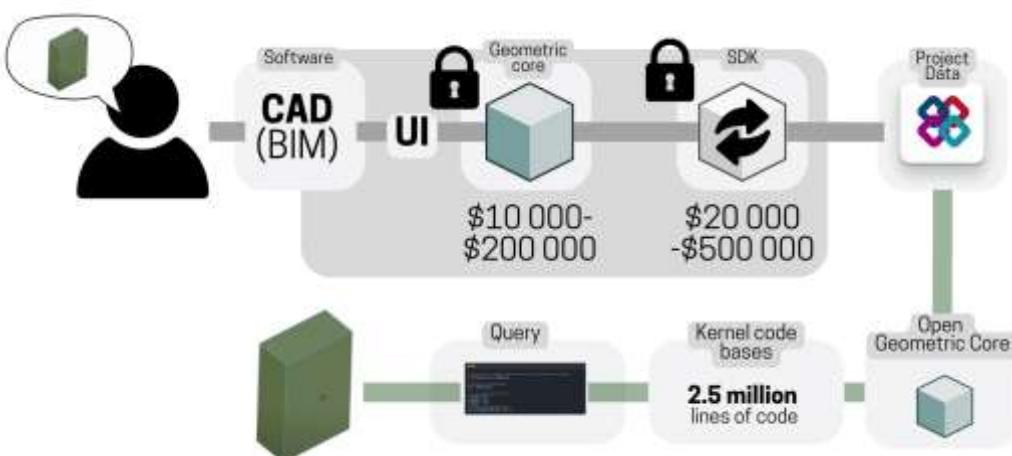
En detaljert oversikt over samspillet mellom mer enn 700 utviklingsteam som har vært involvert i utviklingen av verktøy for å skape og behandle konstruksjonsdata, er presentert i kartet "The Evolution of CAD (BIM)" [116]. [116].

Den åpne formen IFC består av en geometrisk beskrivelse av konstruksjonselementene og en beskrivelse av metainformasjonen. Det brukes ulike metoder for å representere geometri i IFC-format, for eksempel CSG og Swept Solids: Den parametriske representasjonen BREP har imidlertid blitt den ledende standarden for overføring av elementgeometri i IFC-format, ettersom dette formatet støttes ved eksport fra CAD- (BIM-)programmer og gir mulighet for redigering av elementer ved import av IFC tilbake til CAD-programmer.

IFC formatproblem avhengig av den geometriske kjernen

I de fleste tilfeller, når geometrien i IFC er parametrisk definert (BREP), er det umulig å visualisere eller hente ut geometriske egenskaper som volum eller areal for prosjektenheter med bare en IFC-fil, fordi det i dette tilfellet kreves en geometrikjern (figur 6.1-5) for å arbeide med og visualisere geometrien, noe som i utgangspunktet mangler.

Geometry kernel er en programvarekomponent som inneholder grunnleggende algoritmer for opprettning, redigering og analyse av geometriske objekter i CAD (CAD), BIM og andre tekniske applikasjoner. Den er ansvarlig for konstruksjon av 2D- og 3D -geometri, samt for operasjoner på den, for eksempel boolske operasjoner, utjevning, skjæringspunkter, transformasjoner og visualisering.

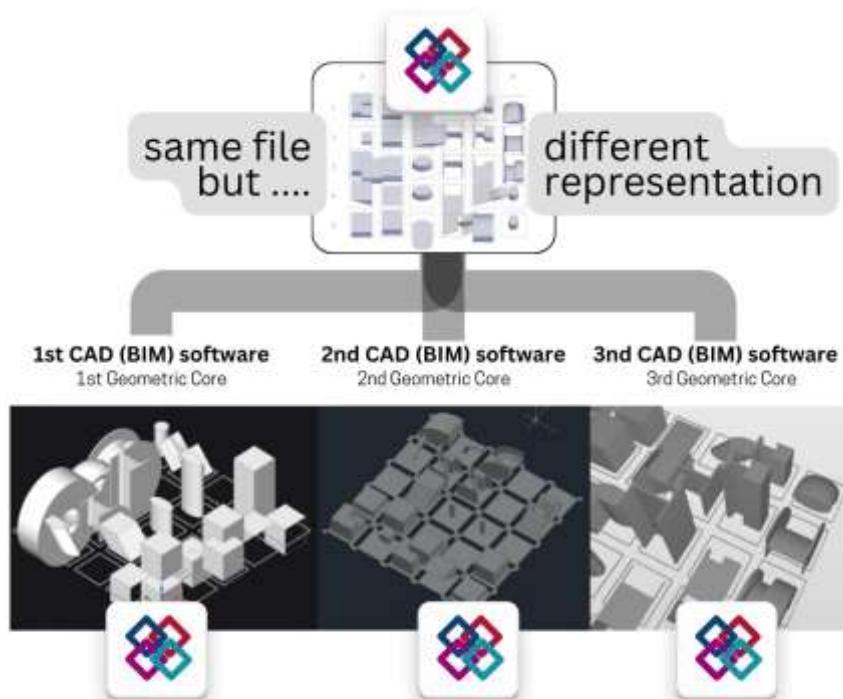


Figur 6.1-5 Opprettelse av geometri gjennom CAD- programvare skjer i dag gjennom proprietære geometrikjerner og SDK-er, som ofte ikke eies av CAD-leverandørene.

Alle CAD-programmer og alle programmer som arbeider med parametriske eller IFC-formater har sin egen eller innkjøpte geometriske kjerner. Og hvis det ikke er noen problemer med primitive elementer i IFC -BREP -format, og i programmer med forskjellige geometriske kjerner kan disse elementene vises på samme måte, men i tillegg til problemer med forskjellige motorer av geometriske kjerner, er det nok elementer som har sine egne særegenheter for riktig visning. Dette problemet er diskutert i detalj i den internasjonale studien "A reference study of IFC software support" publisert i 2019 [117].

De samme standardiserte datasettene gir motstridende resultater, med få felles mønstre, og det er funnet alvorlige problemer med å støtte standarden [IFC], sannsynligvis på grunn av den svært høye kompleksiteten i standarddatamodellen. Standardene har selv en del av skylden for dette, ettersom de ofte etterlater en del detaljer udefinert, med store frihetsgrader og mange tolkningsmuligheter. De tillater høy kompleksitet i organiseringen og lagringen av objekter, noe som ikke bidrar til effektiv universell forståelse, unike implementasjoner og konsistent datamodellering [117].

- Referansestudie av støtte for IFC-programvare, 2021



Source: Reference study of IFC software support: the GeoBIM benchmark 2019 — Part I <https://arxiv.org/pdf/2007.10951>

Fig. 6.1-6 Ulike geometriske kjerner gir forskjellige representasjoner av den samme geometrien beskrevet parametrisk (basert på [117]).

Den korrekte forståelsen av "visse bestemmelser" er tilgjengelig for betalende medlemmer av spesielle organisasjoner som utvikler IFC. Den som ønsker å få tilgang til viktig kunnskap om visse funksjoner i IFC, må derfor forsøke å samarbeide med de store CAD- leverandørene, eller å komme frem til en kvalitativ vurdering av funksjonene ved hjelp av egne undersøkelser

Du snubler over et spørsmål om import og eksport av data via IFC-formatet og spør andre leverandører: "Hvorfor er det slik i IFC-filen informasjon om parametrisk overføring av lokaler? Den åpne spesifikasjonen sier ikke noe om det". Svar fra "mer kunnskapsrike" europeiske leverandører: "Ja, det er ikke sagt, men det er tillatt".

- Fra intervjuet av CAD 2021-utvikleren [118]

IFC beskriver geometrien gjennom parametriske primitiver, men inneholder ikke en innebygd kjerne - denne rollen ivaretas av CAD-programmet, som kompilerer geometrien gjennom geometrikjernen. Geometrikjernen utfører de matematiske beregningene og definerer skjæringspunktene, mens IFC kun leverer dataene for tolkningen. Hvis IFC-en inneholder feilflater, kan forskjellige programmer med forskjellige geometrikjerner enten ignorere dem eller produsere feil, avhengig av kjernen.

For å kunne arbeide med IFC-formatet er det derfor nødvendig å svare på hovedspørsmålet, som det er vanskelig å finne et entydig svar på - hvilket verktøy, med hvilken geometrisk kjerne, skal brukes for å få den samme datakvaliteten som prosjektet opprinnelig hadde i CAD-programmet som IFC ble hentet fra?

Datakvalitsproblemer og kompleksiteten i IFC-formatet tillater ikke direkte bruk av designdata for prosessautomatisering, analyse databehandling, noe som ofte fører utviklere til det uunngåelige behovet for å bruke lukkede CAD-løsninger med "kvalitets"-tilgang til data[63], noe som ble skrevet om av leverandøren selv, som registrerte IFC i 1994 [65].

Alle særegenheter ved mapping og generering av IFC-parametere i geometrikjernen kan bare realiseres av store team av utviklere som har erfaring med å arbeide med geometrikjerner. Derfor er den nåværende praksisen med IFC-formatets særegenheter og kompleksitet først og fremst fordelaktig for CAD-leverandører og har mye til felles med store programvareleverandørers strategi "adoptere, utvide, ødelegge", når standardens økende kompleksitet faktisk skaper barrierer for små markedsaktører [94].

Strategien til store leverandører i en slik strategi kan være tilpasset åpne standarder, legge til egne utvidelser og funksjoner for å skape brukeravhengighet av sine produkter for deretter å drive ut konkurrentene.

IFC-formatet, som er ment å være en universell bro mellom ulike CAD- (BIM-) systemer, fungerer i virkeligheten som en indikator på kompatibilitetsproblemer mellom de geometriske kjernene i ulike CAD-plattformer, på samme måte som STEP-formatet som det opprinnelig ble utviklet fra.

Som et resultat av dette er det i dag mulig for store CAD-leverandører å implementere IFC ontologien fullt ut og av høy kvalitet, og de kan investere betydelige ressurser i å støtte alle entiteter og deres mapping til sin egen interne geometrikjerner, noe som ikke finnes for IFC som standard. Store leverandører har også muligheten til å koordinere tekniske detaljer om funksjoner seg imellom, noe som kanskje ikke er tilgjengelig for selv de mest aktive deltakerne i IFC-formatets utviklingsorganisasjoner.

For små, uavhengige team og åpen kildekode-prosjekter, som forsøker å støtte utviklingen av interoperable formater, blir mangelen på en egen geometrikjernerne et alvorlig problem. Uten en slik kjerne er det praktisk talt umulig å ta hensyn til alle de ulike finessene og nyansene som er forbundet med datautveksling på tvers av plattformer.

Med utviklingen av det parametriske IFC-formatet og det åpne BIM-konseptet har diskusjonene i byggebransjen intensivert seg om hvilken rolle ontologi og semantikk spiller i data- og prosesshåndtering.

Utseende i konstruksjonen av emnet semantikk og ontologi

Takket være ideene om det semantiske internett slutten av 1990-tallet og innsatsen til organisasjoner som var involvert i utviklingen av IFC-formatet, har semantikk og ontologier blitt et av de viktigste elementene i den standardiseringen som diskuteres i byggebransjen på midten av 2020-tallet.

Semantiske teknologier gjør det mulig å forene, standardisere og endre store mengder heterogene data og å gjennomføre komplekse søk.

OWL (Web Ontology Language), representert som RDF-grafer -tripletter (Resource Description Framework) (fig. 6.1-7), brukes til lagring av semantiske data. OWL refererer til grafdatamodeller, som vi har gått nærmere inn på i kapittelet "Datamodeller: datarelasjoner og relasjoner mellom elementer".

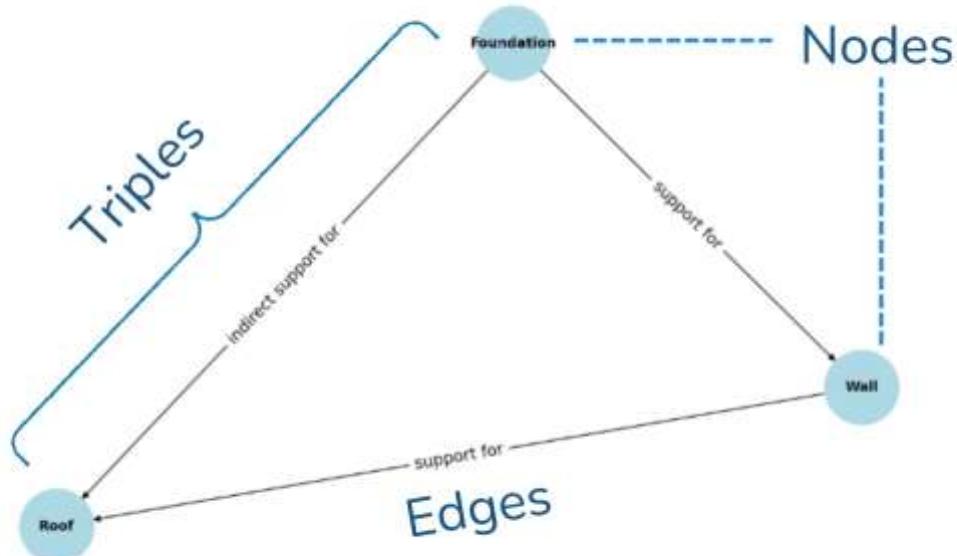


Fig. 6.1-7 RDF-datamodell: Noder, kanter og tripler som illustrerer forholdet mellom byggsteinene.

Teoretisk sett gjør den logiske inferensen til risonere (programvare for automatisk logisk inferens) det mulig å utlede nye utsagn fra ontologier. Hvis for eksempel bygningsontologien registrerer at "et fundament er en støtte for en vegg" og "en vegg er en støtte for et tak" (fig. 6.1-7), kan risoneren automatisk utlede at "et fundament er en støtte for et tak".

En slik mekanisme er nyttig for å optimalisere dataanalysen, fordi den unngår å eksplisitt foreskrive alle avhengigheter. Den skaper imidlertid ikke ny kunnskap, men identifiserer og strukturerer kun allerede kjente fakta.

Semantikk skaper ikke ny mening eller kunnskap i seg selv, og er ikke bedre enn andre datalagrings- og databehandlingsteknologier på dette området. Å representere data fra relasjonsdatabaser som tripleller gjør dem ikke mer meningsfulle. Å erstatte tabeller med grafstrukturer kan være nyttig for å forene datamodeller, gjøre det lettere å finne dem igjen og redigere dem på en sikker måte, men det gjør ikke dataene "smartere" - datamaskinen begynner ikke å forstå innholdet bedre.

Logiske relasjoner i data kan organiseres uten komplekse semantiske teknologier (fig. 6.1-8). Tradisjonelle relasjonsdatabaser (SQL) samt CSV eller XLSX -formatene gjør det mulig å bygge opp lignende avhengigheter. I en kolonnedbatabase kan du for eksempel legge til et felt for "takstøtte" og automatiske knytte taket til fundamentet når du oppretter en vegg. Denne tilnærmingen er implementert uten bruk av RDF, OWL, grafer eller risonere, og er en enkel og effektiv løsning for lagring og analyse av data.

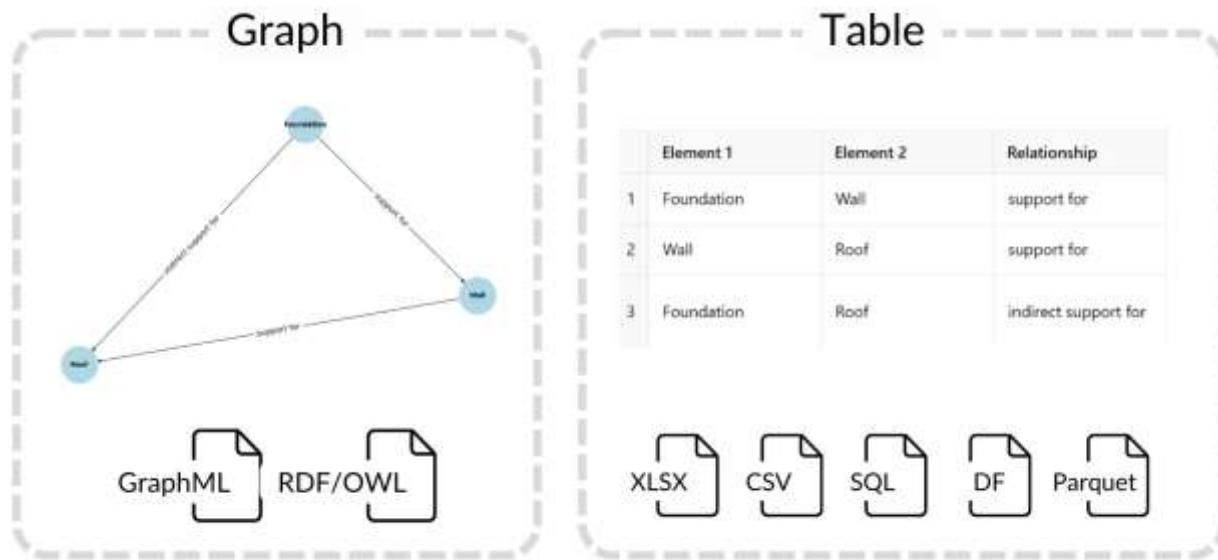


Fig. 6.1-8 Sammenligning av graf- og tabelldatamodeller for representasjon av de samme logiske sammenhengene.

Beslutningen til en rekke store byggfirmaer og IFC-formatutviklingsorganisasjonen [94] om å følge det semantiske webkonseptet, som virket lovende på slutten av 1990-tallet, har hatt en betydelig innvirkning på utviklingen av standarder i byggebransjen.

Paradokset er imidlertid at selve konseptet med det semantiske nettet, som opprinnelig var ment for Internett, ikke har blitt tatt i bruk i særlig grad selv i sitt opprinnelige miljø. Til tross for utviklingen av RDF og OWL har det fullverdige semantiske nettet ikke dukket opp i sin opprinnelige form, og det er allerede

usannsynlig at det vil bli opprettet.

Hvorfor semantiske teknologier ikke lever opp til forventningene i byggebransjen

Andre bransjer har møtt begrensningene i teknologien for å utnytte semantikk. I spillbransjen har forsøk på å beskrive spillobjekter og samspillet mellom dem ved hjelp av ontologier vist seg å være ineffektive på grunn av den høye endringsdynamikken. Derfor har man valgt enklere dataformater som XML og JSON, sammen med algoritmiske løsninger. Situasjonen var den samme i eiendomsbransjen: På grunn av regionale forskjeller i terminologi og hyppige markedsendringer viste det seg å være for komplisert å bruke ontologier, mens enkle databaser og standarder som RETS [119] var bedre i stand til å takle utfordringene ved datautveksling.

Tekniske vanskeligheter, som kompleksiteten i markeringen, arbeidskrevende støtte og lav motivasjon hos utviklerne, bremset innføringen av det semantiske nettet og i andre sektorer av økonomien. RDF (Resource Description Framework) ble ikke en massestandard, og ontologiene viste seg å være for komplekse og økonomisk uforsvarlige.

Den ambisiøse ideen om å skape et globalt semantisk nett ble derfor ikke realisert. Selv om enkelte elementer av teknologien, som ontologier og SPARQL, har funnet veien inn i bedriftsløsninger, er det opprinnelige målet om å skape én helhetlig datastruktur ikke nådd.

Konseptet med et Internett der datamaskiner er i stand til å forstå innholdet, har vist seg å være teknisk vanskelig og kommersielt ulønnsomt. Derfor har selskapene som støttet ideen, etter hvert redusert bruken til enkelte nyttige verktøy, og overlatt RDF og OWL til høyspesialiserte bedriftsbehov snarere enn til Internett som helhet. En analyse av Google Trends (figur 6.1-9) de siste 20 årene tyder på at det semantiske nettet kanskje ikke har flere framtidsutsikter.

Det er ingen grunn til å multiplisere enheter unødig. Hvis det finnes flere logisk konsistente forklaringer på et fenomen som forklarer det like godt, bør man, alt annet likt, foretrekke den enkleste av dem.

barberkniv

Et logisk spørsmål melder seg her: Hvorfor i det hele tatt bruke tripleller, risers og SPARQL i konstruksjonen, når man kan behandle data ved hjelp av populære strukturerede spørninger (SQL, Pandas, Apache®)? I bedriftsapplikasjoner er SQL standarden for arbeid med databaser. SPARQL krever derimot komplekse grafstrukturer og spesialisert programvare, og ifølge trendene i Google vekker det ikke utviklernes interesse.

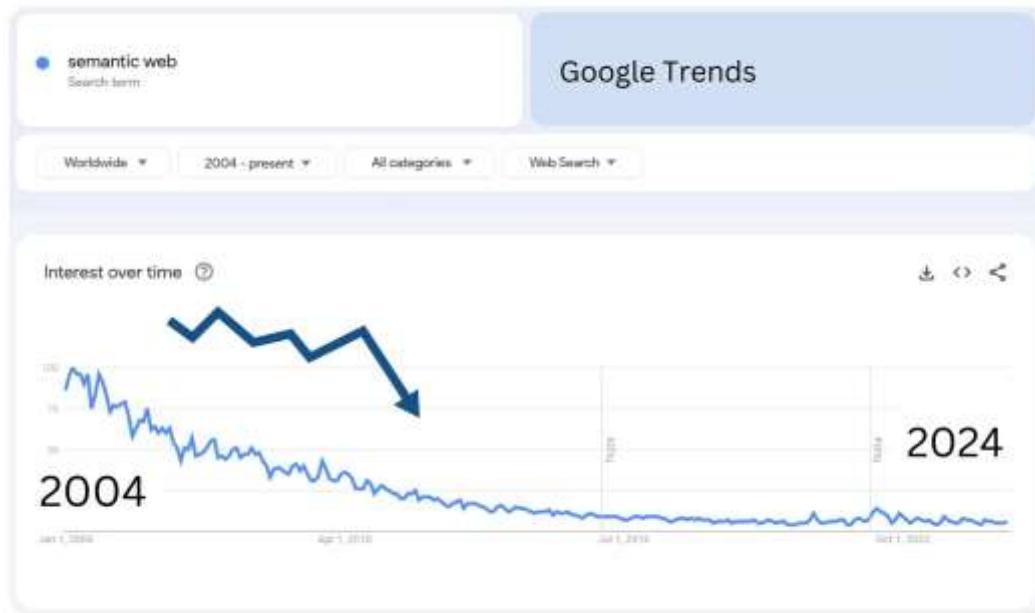


Fig. 6.1-9 Interessen for "semantisk internett"-spørsmål ifølge Googles statistikk.

Grafdatabaser og klassifikasjonstrær kan være nyttige i noen tilfeller, men det er ikke alltid det er grunnlag for å bruke dem i de fleste dagligdagse oppgaver. Derfor er det bare fornuftig å lage kunnskapsgrafer og bruke semantiske webteknologier når det er nødvendig å forene data fra ulike kilder eller trekke komplekse logiske slutninger.

Å gå fra tabeller til grafiske datamodeller forbedrer søket og forenkler informasjonsflyten, men det gjør ikke dataene mer meningsfulle for maskiner. Spørsmålet er ikke om semantiske teknologier bør brukes, men hvor de virkelig gjør en forskjell. Før du implementerer ontologi, semantikk og graf databaser i din bedrift, bør du finne ut hvilke bedrifter som allerede bruker disse teknologiene med suksess, og hvor de har mislyktes.

Til tross for ambisiøse forventninger ble semantiske teknologier aldri en universell løsning for strukturering av data i byggebransjen. I praksis har disse teknologiene ikke ført til en universell løsning, men bare bidratt til ny kompleksitet, noe som gjenspeiler de urealiserte ambisjonene for det semantiske Internett-konseptet, der forventningene langt overgikk virkeligheten.



Fig. 6.1-10 Geometri og informasjon i byggeprosesser: fra komplekse CAD- og BIM- systemer til forenklet data for analyse.

Mens IT-industrien har fått kompensert for det semantiske nettets fiaskoer med fremveksten av ny teknologi (stordata, IoT, maskinlæring, AR/VR), har byggebransjen ingen slike muligheter.

I tillegg til utfordringene med å bruke konsepter for å kommunisere datarelasjoner mellom prosjektelelementer, gjenstår et grunnleggende problem - selve tilgjengeligheten av disse dataene. Byggebransjen domineres fortsatt av lukkede systemer, noe som gjør det vanskelig å jobbe med data, dele informasjon og forbedre prosesseffektiviteten.

Det er dataenes lukkede natur som er i ferd med å bli en av de viktigste barrierene som hindrer utviklingen av digitale løsninger i byggebransjen. I motsetning til IT-bransjen, der åpne og harmoniserte dataformater har blitt standard, bruker hver programvare sitt eget format i CAD-sektoren (BIM), noe som skaper lukkede økosystemer og kunstig begrenser brukerne.



KAPITTEL 6.2.

LUKKEDE PROSJEKTFORMATER OG PROBLEMER MED INTEROPERABILITET

Lukkede data og fallende produktivitet: CAD-bransjens blindvei (BIM)

CAD -systemenes proprietære natur har ført til at hvert program har sitt eget unike dataformat, som enten er lukket og utilgjengelig fra utsiden - RVT, PLN, DWG, NDW, NWD, SKP, eller er tilgjengelig i semistrukturert form gjennom en ganske kompleks konverteringsprosess - JSON, XML (CPIXML), IFC, STEP og ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, osv..

De ulike dataformatene som de samme dataene om de samme prosjektene kan lagres i, har ikke bare ulik struktur, men inkluderer også ulike versjoner av den interne markeringen, noe utviklere må ta hensyn til for å sikre applikasjonskompatibilitet. For eksempel vil et CAD-format fra 2025 kunne åpnes i et CAD-program fra 2026, men det samme prosjektet vil aldri kunne åpnes i alle versjoner av CAD-programmet som kan ha vært tilgjengelige før 2025.

Ved å ikke gi direkte tilgang til databaser, skaper programvareleverandører i byggebransjen ofte sine egne unike formater og verktøy som en fagperson (prosjekterende eller dataansvarlig) må bruke for å få tilgang til, importere og eksportere data.

Som en konsekvens av dette øker leverandørene av grunnleggende CAD (BIM) og relaterte løsninger (f.eks. ERP/PMIS)) stadig prisene for å bruke produktene, og vanlige brukere tvinges til å betale en "provisjon" på hvert trinn i dataoverføringen av formater [63]: for å koble til, importere, eksportere og arbeide med data som brukerne selv har skapt.

Kostnaden for å få tilgang til data i skylagring fra populære CAD - (BIM-) produkter vil nå 1 dollar per transaksjon i 2025 [120], og abonnementer på ERP -produkter for mellomstore bedrifter når opp i fem- og sekssifrede beløp per år [121].

Essensen i moderne byggprogramvare er at det ikke er automatisering eller økt effektivitet, men ingeniørenes evne til å forstå en bestemt høyspesialisert programvare som påvirker kvaliteten og kostnadene ved databehandling av byggeprosjekter, samt fortjenesten og den langsiktige overlevelsen til selskaper som gjennomfører byggeprosjekter.

Mangelen på tilgang til databaser CAD -systemer som brukes i dusinvis av andre systemer og hundrevis av prosesser [63], og den påfølgende mangelen på kvalitetskommunikasjon mellom individuelle spesialister har ført til at byggebransjen har status som en av de mest ineffektive sektorene i økonomien når det gjelder produktivitet [44].

I løpet av de siste 20 årene med CAD- (BIM-) designapplikasjoner, fremveksten av nye systemer (ERP), nye konstruksjonsteknologier og materialer, har produktiviteten i hele byggebransjen sunket med 20% (figur 2.2-1), mens den samlede produktiviteten i alle sektorer av økonomien som ikke har store problemer med å få tilgang til databaser og markedsføringslignende BIM-konsepter, har økt med 70% (96% i

produksjonsindustrien) [122].

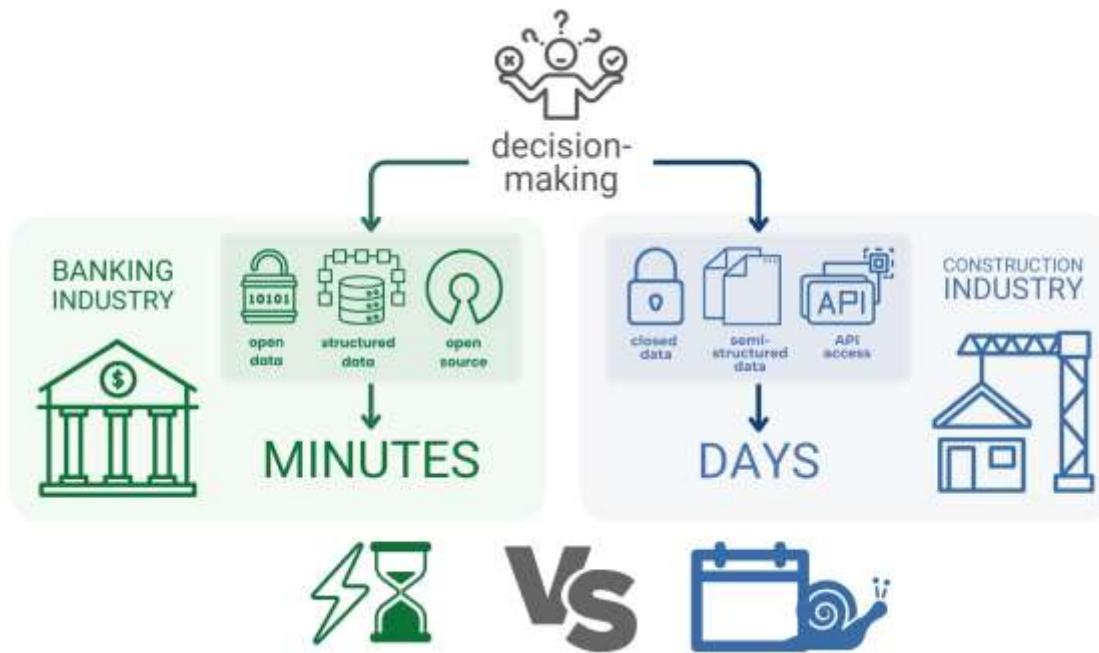


Fig. 6.2-1 På grunn av isolasjonen og kompleksiteten i prosjektdataene som dusinvis av avdelinger og hundrevis av prosesser er avhengige av i byggebransjen, er beslutningshastigheten flere ganger langsmmere enn i andre bransjer.

Det finnes imidlertid også enkelte eksempler på alternative tilnæringer til å skape interoperabilitet mellom DAK-løsninger. Europas største byggefirma viser med SCOPE-prosjektet [123], som ble startet i 2018, hvordan det er mulig å gå utover den klassiske logikken i CAD- (BIM)-systemer. I stedet for å prøve å underlegge seg IFC eller basere seg på proprietære geometrikjerner, bruker SCOPE-utviklere API-er og SDK-er reverse engineering til å hente ut data fra ulike DAK-programmer, konvertere dem til nøytrale formater som OBJ eller CPIXML basert på den eneste Open Source geometrikjernen OCCT, og videre bruke dem i hundrevis av forretningsprosesser i bygg- og designfirmaer. Til tross for at ideen er progressiv, møter slike prosjekter begrensningene og kompleksiteten til gratis geometriske kjerner, og de forblir fortsatt en del av lukkede økosystemer i ett selskap som reproducerer logikken til løsninger med én leverandør.

På grunn av begrensningene i lukkede systemer og forskjeller i dataformater, samt mangelen på effektive verktøy for å forene dem, står bedrifter som må arbeide med CAD-formater, overfor en opphopning av betydelige mengder data med varierende grad av struktur og lukkethet. Disse dataene blir ikke brukt på riktig måte og forsvinner i arkiver, der de forblir glemt og ubrukt for alltid.

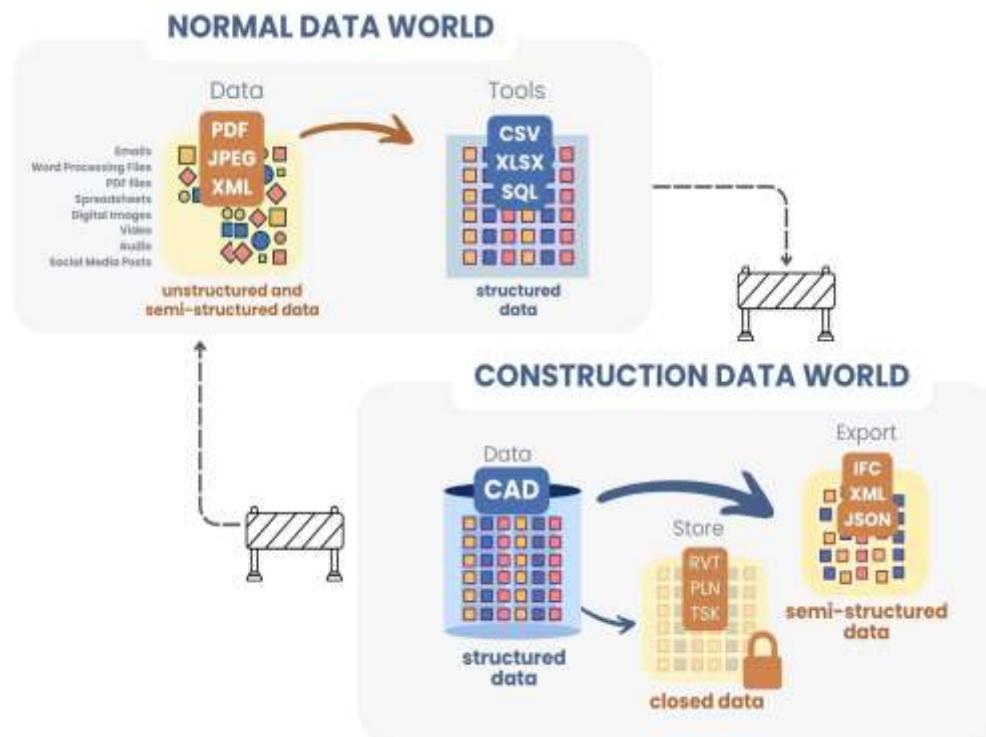
Data som er innhentet gjennom en betydelig innsats i designfasen, blir utilgjengelige for videre bruk på grunn av sin kompleksitet og lukkede natur.

I løpet av de siste 30 årene har utviklere i byggebransjen derfor blitt tvunget til å møte det samme problemet om og om igjen: Hvert nye lukkede format eller hver nye proprietære løsning skaper et behov for å integrere med eksisterende åpne og lukkede DAK-systemer. Disse stadige forsøkene på å sikre interoperabilitet mellom ulike DAK- og BIM-løsninger bidrar bare til å komplisere dataøkosystemet, i stedet for å bidra til

forenkling og standardisering.

Myten om interoperabilitet mellom CAD-systemer

Mens den viktigste retningen for utviklingen av interoperabilitet i DAK-miljøet på midten av 1990-tallet var å bryte med det proprietære DWG-formatet - noe som kulminerte i seieren til Open DWG-alliansen [75] og den faktiske åpningen av det mest populære tegneformatet for hele byggebransjen - har fokuset skiftet på midten av 2020-tallet. En ny trend er i ferd med å skyte fart i byggebransjen: En rekke utviklingsteam fokuserer på å skape såkalte "broer" mellom lukkede CAD-systemer (lukket BIM), IFC-formatet og åpne løsninger (åpen BIM). De fleste av disse initiativene er basert på bruk av IFC-formatet og OCCT-geometrikjernen, som utgjør en teknisk bro mellom ulike plattformer. Denne tilnærmingen anses som en lovende retning som kan forbedre datautvekslingen og interoperabiliteten mellom programvareverktøy betydelig.

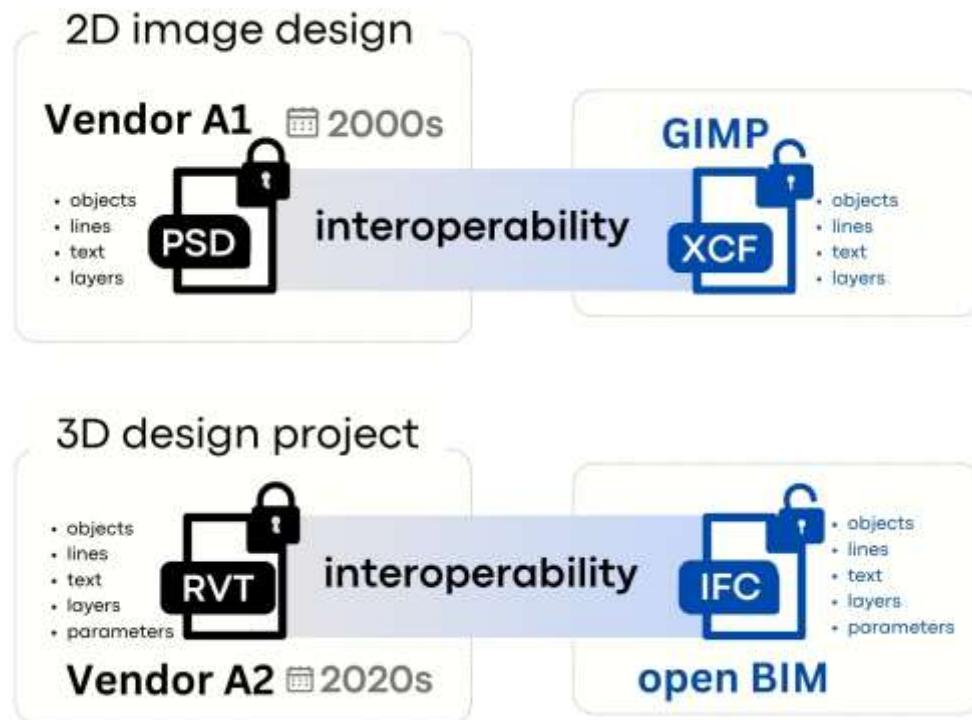


Figur 6.2-2 Mens andre bransjer arbeider med åpne data, må byggebransjen arbeide med lukkede eller løst strukturerede DAK-formater (BIM).

En slik tilnærming har historiske paralleller. På 2000-tallet prøvde utviklere å overvinne dominansen til den største leverandøren av grafiske redigeringsprogrammer (2D world), og forsøkte å skape en sømløs integrasjon mellom sin proprietære løsning og gratis Open Source - et alternativ til GIMP (fig. 6.2-3). Den gang, som i dag, handlet det om å forsøke å bygge bro mellom lukkede og åpne systemer og samtidig bevare komplekse parametere, lag og intern programvarelogikk.

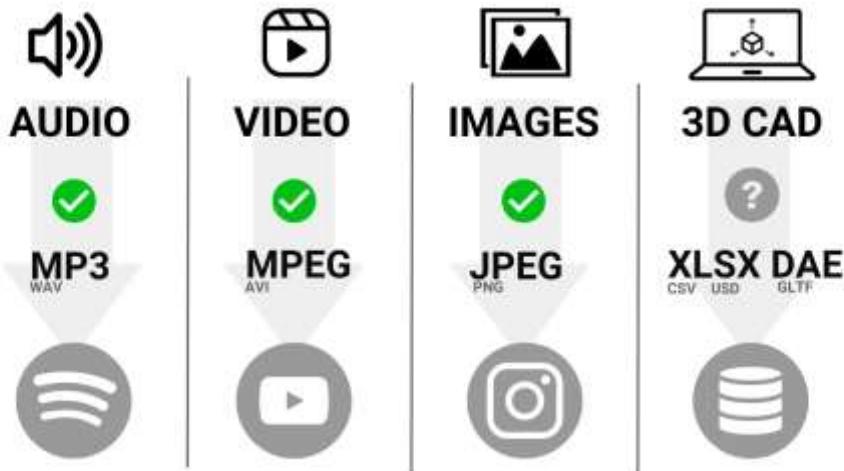
Brukerne var imidlertid på jakt etter enkle løsninger - flate, åpne data uten overdreven kompleksitet i form av lag og programparametere (en analog til den geometriske kjernen i CAD). Brukerne ønsket enkle og åpne

dataformater uten overdreven logikk. JPEG, PNG og GIF ble slike formater innen grafikk. I dag brukes de i sosiale nettverk, på nettsteder og i applikasjoner - de er enkle å behandle og tolke, uavhengig av plattform eller programvareleverandør.



Figur 6.2-3 Interoperabiliteten mellom dataformater i bygg- og anleggsbransjen kan sammenlignes med forsøkene på å slå sammen en populær leverandørs proprietære produkt og GIMP med åpen kildekode på 2000-tallet.

Resultatet er at nesten ingen i bildebransjen i dag bruker lukkede formater som PSD eller åpne XCF til applikasjoner, sosiale nettverk som Facebook og Instagram, eller som innhold på nettsteder. I stedet bruker de fleste flate og åpne JPEG-, PNG- og GIF-formater for brukervennlighet og bred kompatibilitet. Åpne formater som JPEG og PNG har blitt standarden for bildedeling på grunn av sin allsidighet og brede støtte, noe som gjør dem enkle å bruke på en rekke ulike plattformer. En lignende overgang kan observeres i andre utvekslingsformater, som video og lyd, der universelle formater som MPEG og MP3 fremheves for sin komprimeringseffektivitet og brede kompatibilitet. En slik utvikling mot standardisering har gjort det enklere å dele og spille av innhold og informasjon, slik at det er tilgjengelig for alle brukere på flere plattformer (figur 6.2-4).



Figur 6.2-4 Forenkede formater uten kompliserte redigeringsfunksjoner har blitt populære for deling og bruk av data.

Lignende prosesser forekommer i 3D modellering. Enkle og åpne formater som USD, OBJ, glTF, DAE, DXF, SQL og XLSX brukes i økende grad i prosjekter for datautveksling utenfor CAD-miljøet (BIM). Disse formatene lagrer all nødvendig informasjon, inkludert geometri og metadata, uten at det er nødvendig å bruke en kompleks BREP-struktur, geometrikjerner eller leverandørspesifikke interne klassifikatorer. Proprietære formater som NWC, SVF, SVF2, CPIXML og CP2 som leveres av ledende programvareleverandører, utfører også lignende funksjoner, men forblir lukkede, i motsetning til åpne standarder.

Det er bemerkelsesverdig (og verdt å minne om igjen, som allerede nevnt i forrige kapittel) at denne ideen - avisningen av mellomliggende nøytrale og parametriske formater som IGES, STEP og IFC - ble støttet allerede i 2000 av den store CAD-leverandøren som laget BIM Whitepaper og registrerte IFC-formatet i 1994. I Whitepaper 2000 "Integrated Design and Manufacturing" [65] understreker CAD-leverandøren viktigheten av å ha direkte tilgang til CAD-databasen i programvaremiljøet, uten behov for å bruke mellomliggende oversettere og parametriske formater, for å opprettholde informasjonens fullstendighet og nøyaktighet.

Byggebransjen har ennå ikke blitt enige om verktøy for å få tilgang til CAD-databaser eller tvungen reverse engineering, eller om å innføre et felles, forenket dataformat for bruk utenfor CAD-plattformer (BIM). For eksempel bruker mange store selskaper i Sentral-Europa og tysktalende regioner i bygg- og anleggsbransjen formatet CPIXML i sine ERP-systemer [121]. Dette proprietære formatet, som er en slags XML, kombinerer CAD (BIM)-prosjektdata, inkludert geometriske data og metadata, i en enkelt organisert og forenket struktur. Store byggfirmaer skaper også egne nye formater og systemer, som i SCOPE-prosjektet, som vi diskuterte i forrige kapittel

Den lukkede logikken i parametriske CAD-formater eller komplekse parametriske filer IFC (STEP) er overflødig i de fleste forretningsprosesser. Brukerne er ute etter forenklede og flate formater som USD, CPIXML, XML & OBJ, DXF, SQLite, DAE & XLSX, som inneholder all nødvendig elementinformasjon, men som ikke er belastet med overflødig BREP-geometri, avhengighet av geometrikjerner og interne klassifikasjoner av spesifikke CAD og BIM -produkter (fig. 6.2-5).

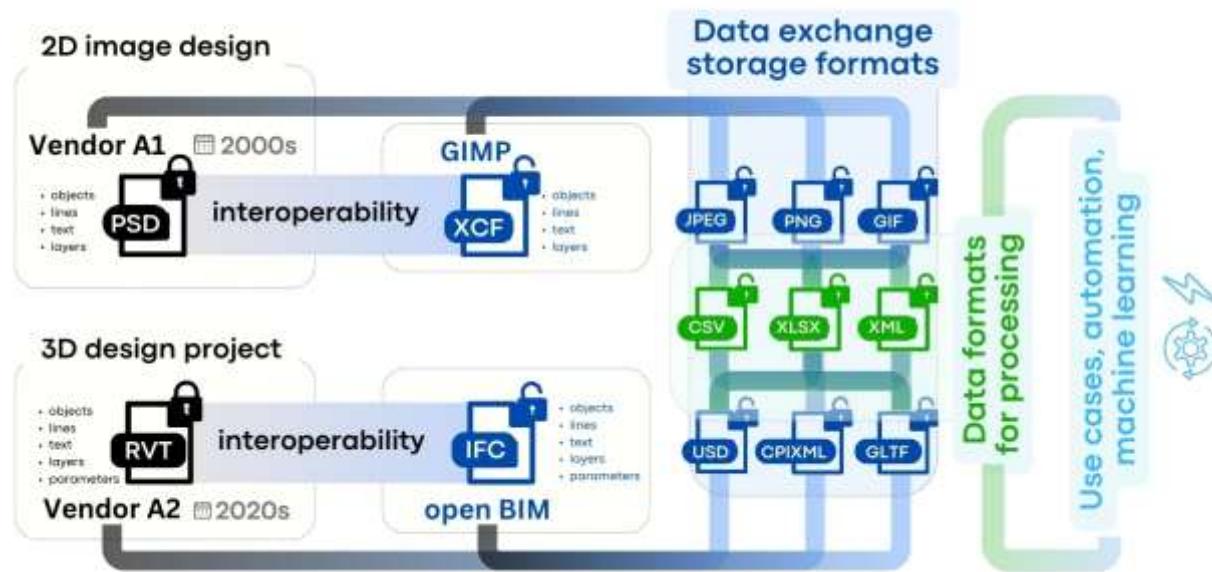


Fig. 6.2-5 For de fleste bruksområder velger brukerne de enkleste formatene som er uavhengige av leverandørprogrammer.

Fremveksten av flate bildeformater som JPEG, PNG og GIF, som er frigjort fra den overflødige logikken i leverandørenes interne motorer, har satt fart i utviklingen av tusenvis av interoperable løsninger for behandling og utnyttelse av grafikk. Dette har ført til fremveksten av applikasjoner som spenner fra retusjerings- og filtreringsverktøy til sosiale nettverk som Instagram, Snapchat og Canva, der disse forenklede dataene kan utnyttes uten å være bundet til en bestemt programvareutvikler.

Standardisering og forenkling av design-CAD-formater vil stimulere fremveksten av mange nye brukervennlige og uavhengige verktøy for arbeid med byggeprosjekter.

Ved å gå bort fra den komplekse logikken i leverandørapplikasjoner som er knyttet til lukkede geometrikjerner, og over til universelle, åpne formater basert på biblioteker med forenklede elementer, skapes det forutsetninger for en mer fleksibel, transparent og effektiv datahåndtering. Dette gir også tilgang til informasjon for alle parter som er involvert i byggeprosessen - fra designere til kunder og vedlikeholdstjenester.

Likevel er det høyst sannsynlig at leverandørene av CAD i løpet av de kommende årene vil forsøke å flytte debatten om interoperabilitet og tilgang til CAD-databaser igjen. Det vil allerede dreie seg om "nye" konsepter - som granulære data, intelligente grafer, "fødererte modeller", digitale tvillinger i skylagre - samt opprettelsen av bransjeallianser og standarder som fortsetter utviklingen av BIM og åpen BIM. Til tross for den tiltalende terminologien kan slike initiativer nok en gang bli verktøy for å holde brukerne innenfor proprietære økosystemer. Et eksempel er den aktive markedsføringen av USD (Universal Scene Description)-formatet som den "nye standarden" for plattformovergripende CAD (BIM)-samarbeid fra 2023.

Gå til USD og detaljerte data m

Fremveksten av AOUSD-alliansen [124] i 2023 markerer en viktig vending i byggebransjen. Vi er vitne til begynnelsen på en ny virkelighet, formet av CAD-leverandører, når det gjelder håndtering av konstruksjonsdata gjennom flere viktige endringer. Den første store endringen gjelder oppfatningen av CAD-data. Fagfolk som er involvert i de tidlige stadiene av den konseptuelle prosjekteringen, innser i økende grad at det å skape et design i et DAK-miljø bare er et utgangspunkt. Dataene som genereres i løpet av designprosessen, blir etter hvert grunnlaget for analyse, drift og administrasjon av objekter. Det betyr at de må være tilgjengelige og brukbare i systemer som går ut over tradisjonelle DAK-verktøy.

Parallelt skjer det en revolusjon i tilnærmingen hos ledende utviklere. Bransjens ledende CAD-leverandør, skaperen av BIM-konseptet og IFC-formatet, tar en uventet vending i sin strategi. Fra 2023 går selskapet bort fra den tradisjonelle lagringen av data i separate filer, fokuserer på å arbeide med granulære (normaliserte og strukturerte) data og går over til en datasentrisk tilnærming [125].

Leverandørene følger de historiske trendene i andre bransjer: De fleste brukerne trenger ikke lukkede CAD-formater (som PSD) eller komplekse parametriske IFC-filer (som GIMP med laglogikk). De trenger enkle objektbilder som kan brukes i CAFM (Construction Instagram), ERP (Facebook) og tusenvis av andre prosesser fylt med Excel regneark og PDF dokumenter.

Dagens trender i byggebransjen kan føre til en gradvis overgang fra parametriske og komplekse formater til fordel for de mer universelle og uavhengige formatene USD, GLTF, DAE, OBJ (med metainformasjon både i hybriden og i separate strukturerte eller løst strukturerte formater). Historiske ledere, inkludert store designselskaper som en gang aktivt promoterte IFC på midten av 1990-tallet, fremmer nå åpent det nye USD-formatet [93], og legger vekt på dets enkelhet og allsidighet (fig. 6.2-6). Massebruken av USD i produkter, GLTF-kompatibiliteten og den aktive integreringen i verktøy som Blender, Unreal Engine og Omniverse viser potensialet for et nytt paradigme for arbeid med data. Sammen med populariteten til lokaliserte løsninger som det europeiske flate USD-formatet CPIXML, som brukes i populære europeiske ERP-systemer, kan potensielt styrke USDs posisjon i Sentral-Europa. Organisasjoner som er involvert i utviklingen av IFC-formatet, er allerede i ferd med å tilpasse strategien sin til USD [126], noe som bare bekrefter at dette skiftet er uunngåelig.

Technical Specifications				Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility	
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data	
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.	
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON	
Export/import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products	
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it.	
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse	
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms	
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools	
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions	
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications	
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure	
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations	
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement	

Figur 6.2-6 Sammenligning av tekniske spesifikasjoner for IFC og USD-formatene.

På denne bakgrunnen har USD potensial til å bli de facto-standarden, og lover å overvinne mange av dagens begrensninger, først og fremst knyttet til kompleksiteten i eksisterende CAD - (BIM-) formater og avhengigheten av deres tolkning av geometriske kjerner.

I stedet for parametriske og komplekse CAD-formater og IFC - vil forenklede dataformater som USD, glTF, DAE, OBJ med elementmetainformasjon i CSV, XLSX, JSON, XML få en plass i byggebransjen på grunn av deres enkelhet og fleksibilitet.

Ved første øyekast ser dagens endringer i byggebransjen ut som et teknologisk gjennombrudd i forbindelse med overgangen fra den utdaterte IFC til den mer moderne USD. Det er imidlertid verdt å tenke på at den samme CAD-leverandøren som utviklet IFC, allerede i 2000 skrev om problemene med IFC og behovet for

tilgang til databasen [65], og som nå aktivt fremmer overgangen til en ny standard - USD.

Bak nok en fasade av "åpne data", USD og "nye" konsepter for detaljert datahåndtering gjennom skybaserte applikasjoner som CAD -leverandørene begynner å markedsføre, kan det skjule seg en intensjon om å monopolisere prosjektdatahåndteringen, der brukerne befinner seg i en situasjon der valg av format er mer knyttet til bedriftens interesser enn til reelle behov.

En analyse av nøkkelfakta [93] viser at hovedformålet med disse endringene i mindre grad handler om brukervennlighet, men først og fremst om å opprettholde kontroll over økosystemer og datastrømmer til fordel for leverandører som i løpet av 40 år aldri har vært i stand til å gi tilgang til CAD-databaser.

Kanskje er det nå på tide at selskapene slutter å vente på nye konsepter fra programvareleverandørene og heller fokuserer på egenutvikling i datasentrisk retning. Etter å ha frigjort seg fra problemer med datatilgang ved hjelp av verktøy for omvendt data teknikk, vil bransjen selv kunne gå over til moderne, gratis og praktiske verktøy for arbeid med og analyse av data uten å påtvinge seg nye konsepter.

CAD (BIM) Maturity Levels: From Stage 0 to Structured Data

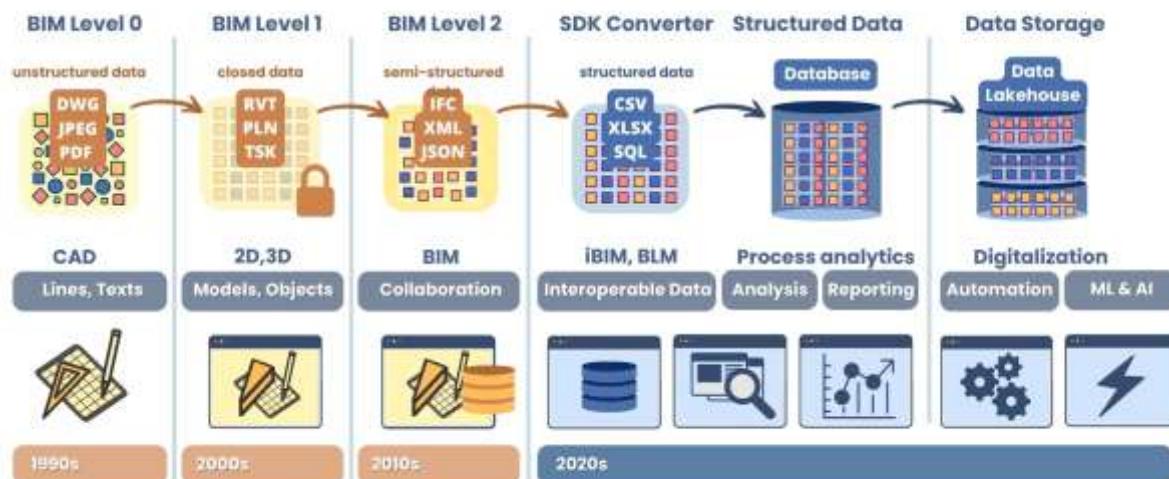


Fig. 6.2-7 CAD-modenhetsnivå (BIM): fra ustukturerte data til strukturerte data og datalagre.

Tilgang til databaser, åpne data og formater vil uunngåelig bli standard i byggebransjen, uavhengig av leverandørenes forsøk på å forsinke prosessen - det er bare et spørsmål om tid (fig. 6.2-7). Tempoet i denne overgangen kan øke betraktelig hvis flere og flere fagfolk blir kjent med åpne formater, databaseverktøy og tilgjengelige SDK-er for omvendt konstruksjon, som gir direkte tilgang til CAD-data - systems [92].

Fremtiden ligger i åpne, enhetlige og analytisk tilgjengelige data. For å unngå å være avhengig av leverandørløsninger og å bli tatt som gissel i lukkede økosystemer, må bygg- og anleggsbedrifter før eller siden satse på åpenhet og uavhengighet, og velge formater og løsninger som gir full kontroll over dataene.

Dataene som skapes i byggebransjen i dag, vil være en nøkkelressurs for forretningsbeslutninger i fremtiden. De vil fungere som det strategiske "drivstoffet" som driver utviklingen og effektiviteten til bygg- og anleggsbedriftene. Byggebransjens fremtid ligger i evnen til å arbeide med data, ikke i valg av dataformater eller modeller.

For å forstå forskjellen mellom de åpne formatene USD, glTF, DAE, OBJ og proprietære parametriske CAD-formater, er det viktig å se nærmere på et av de mest komplekse og sentrale dataelementene i visualisering og konstruksjonsberegninger - geometri og genereringsprosessene for denne. Og for å forstå hvordan geometriske data danner grunnlaget for analyser og beregninger i bygg- og anleggsbransjen, er det nødvendig å se nærmere på mekanismene for generering, transformasjon og lagring av geometri.



KAPITTEL 6.3.

GEOMETRI I BYGG OG ANLEGG: FRA LINJER TIL KUBIKKMETER

Når linjer blir til penger eller hvorfor byggherrer trenger geometri

Geometri i konstruksjonen er ikke bare en visualisering, men også grunnlaget for nøyaktige kvantitative beregninger. I prosjektmodellen supplerer geometrien listene over elementparametere (fig. 3.1-16) med viktige volumetriske egenskaper som lengde, areal og volum. Disse volumetriske parameterverdiene beregnes automatisk ved hjelp av geometrikjernen og er utgangspunktet for estimatorer, tidsplaner og ressursmodeller. Som vi allerede har diskutert i del 5 av denne boken og i kapittelet "Kostnadsberegnning og estimering byggeprosjekter", er det de volumetriske parametrene til grupper av objekter fra CAD -modeller som danner grunnlaget for moderne ERP-, PMIS -systemer/ Geometri spiller en grunnleggende rolle ikke bare i prosjekteringsfasen, men også i prosjektgjennomføringsstyring, tidsplanstyring, budsjettering og drift. På samme måte som nøyaktigheten av et prosjekt var avhengig av lengdemål som alen og alen for tusenvis av år siden, da de egyptiske pyramidene ble bygget, har nøyaktigheten av geometritolkningen i CAD -programmer i dag direkte innvirkning på resultatet: fra budsjett tidsfrister til valg av entreprenør og leveringslogistikk.

I et svært konkurransesatt og budsjettbegrenset miljø blir nøyaktigheten av volumetriske beregninger, som er direkte avhengig av geometrien, en overlevelsесfaktor. Moderne ERP -systemer er direkte avhengige av korrekte volumetriske egenskaper hentet fra CAD - og BIM -modeller. Derfor er nøyaktig geometrisk beskrivelse av elementer ikke bare en visualisering, men et nøkkelverktøy for å styre kostnader og byggetid.

Historisk sett har geometri vært det viktigste språket for ingeniørkommunikasjon. Fra streker på papyrus til digitale modeller har tegninger og geometriske representasjoner fungert som et middel til å utveksle informasjon mellom designere, formenn og kalkulatorer. Før datamaskinene kom, ble beregningene gjort manuelt ved hjelp av linjaler og vinkelmålere. I dag er denne oppgaven automatisert takket være volumetrisk modellering: De geometriske kjernene i CAD-programvaren konverterer linjer og punkter til tredimensjonale legemer som alle de nødvendige egenskapene automatisk trekkes ut fra.

I CAD -programmer opprettes geometriske elementer for beregninger gjennom brukergrensesnittet til CAD- (BIM-) -programmer. For å omdanne punkter og linjer til volumetriske legemer brukes den geometriske kjernen, som oppfyller nøkkelloppgaven - omdanning av geometri til volumetriske modeller, hvorfra elementets volumetriske egenskaper beregnes automatisk etter tilnærming.

Fra linjer til volumer: Hvordan areal og volum blir til data

I ingeniørpraksis beregnes volumer og arealer ut fra geometriske flater som beskrives analytisk eller ved hjelp av parametriske modeller som NURBS (nonuniform rational B-splines) innenfor BREP-rammeverket (boundary element representation).

NURBS (Non-Uniform Rational B-Splines) er en matematisk måte å beskrive kurver og overflater på, mens BREP er et rammeverk for å beskrive hele den tredimensjonale geometrien til et objekt, inkludert grensene, som kan defineres ved hjelp av NURBS.

Til tross for nøyaktigheten til BREP og NURBS, krever de kraftige beregningsressurser og komplekse algoritmer. Det er imidlertid ofte vanskelig å gjøre direkte beregninger ut fra slike matematisk nøyaktige beskrivelser, så i praksis brukes nesten alltid tessellering - omforming av flater til et rutenett av trekanner - noe som forenkler de påfølgende beregningene. Tessellering er oppdelingen av en kompleks overflate i trekanner eller polygoner. I CAD /CAE -miljøer brukes denne metoden til visualisering, volumberegninger, kollisjonssøk, eksport til formater som MESH og kollisjonsanalyse. Et eksempel fra naturen er en bikube, der en kompleks form brytes ned til et regelmessig rutenett (fig. 6.3-1).

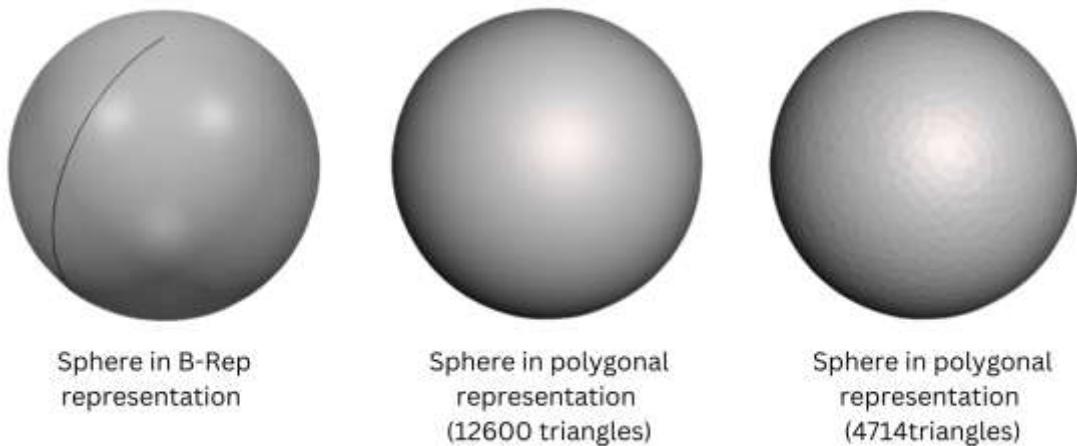


Fig. 6.3-1 Samme kule i parametrisk beskrivelse BREP og polygonal representasjon med forskjellig antall trekanner.

BREP (NURBS), som brukes i CAD, er ikke en grunnleggende modell for geometri. Den ble laget som et praktisk verktøy for å representer sirkler og rasjonelle splines og for å minimere lagring av geometridata. Den har imidlertid sine begrensninger - for eksempel at den ikke er i stand til å beskrive sinuskurven som ligger til grunn for spiralfomede linjer og flater, og at det er nødvendig å bruke komplekse geometriske kjerner.

Triangulære nett og tessellering av parametriske former kjennetegnes derimot av enkelhet, effektiv bruk av minne og muligheten til å behandle store datamengder (fig. 6.3-2). Disse fordelene gjør det mulig å klare seg uten komplekse og kostbare geometriske kjerner, og de titalls millioner linjene med kode som er innebygd i dem, når geometriske former skal beregnes.

I de fleste tilfeller spiller det ingen rolle hvor nøyaktig de volumetriske egenskapene er definert - gjennom parametriske modeller (BREP, IFC) eller gjennom polygoner (USD, glTF, DAE, OBJ). Geometrien forblir en tilnærming: enten det er gjennom NURBS eller MESH, er det alltid en omrentlig beskrivelse av formen.

Geometri definert som polygoner eller BREP (NURBS) er til en viss grad bare en måte å tilnærme seg med en tilnærmet beskrivelse av en kontinuerlig form. På samme måte som Fresnel-integraler ikke har noe eksakt analytisk uttrykk, er diskretisering av geometri gjennom polygoner eller NURBS alltid en tilnærming, akkurat som triangulær MESH.

Parametrisk geometri i BREP-format er nødvendig hovedsakelig der det er viktig med minimal datastørrelse og det er mulig å bruke ressurskrevende og dyre geometrikjerner for behandling og visning. Dette er ofte karakteristisk for utviklere av CAD-programmer, som for dette formålet bruker geometriske kjerner fra MCAD-leverandører i sine produkter. I dette tilfellet, selv i disse programmene, blir BREP-modeller i prosessen med tessellering for visualisering og beregninger ofte konvertert til trekant (på samme måte som PSD-filer blir forenklet til JPEG).

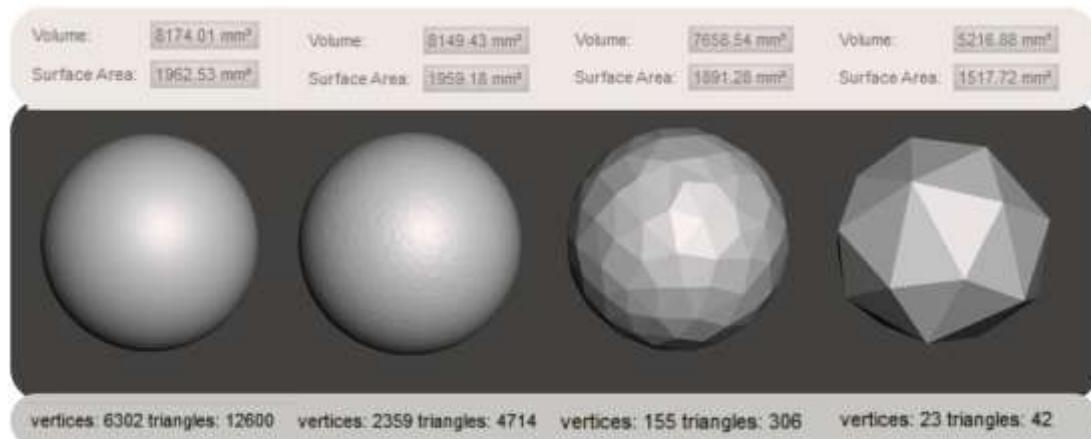


Fig. 6.3-2 Forskjell i volumetriske egenskaper i figurer med forskjellig antall polygoner.

Polygonal MESH, så vel som parametrisk BREP, har sine egne fordeler og begrensninger, men målet er det samme - å beskrive geometrien med hensyn til brukerens oppgaver. Til syvende og sist avhenger nøyaktigheten til en geometrisk modell ikke bare av representasjonsmetoden, men også av kravene til en bestemt oppgave.

I de fleste konstruksjonsproblemer kan behovet for parametrisk geometri og komplekse geometriske kjerner være overflødig.

I hver enkelt automatiseringsoppgave er det verdt å vurdere om betydningen av parametrisk geometri overdrives av CAD-utviklere som er interessert i å markedsføre og selge sine egen programvareprodukter.

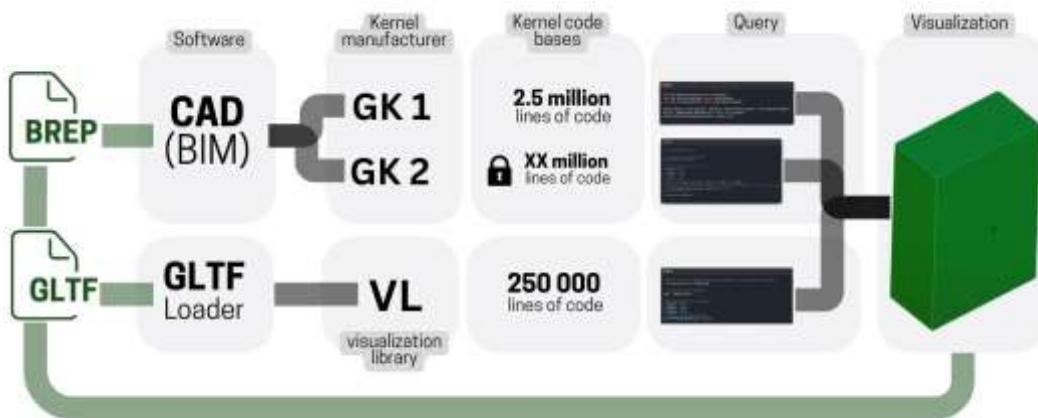
Overgang til MESH, USD og polygoner: bruk av tessellering for geometri

I byggebransjen er det viktig å tilstrebe uavhengighet fra spesifikke DAK-editorer og geometrikjerner når man streamer, utvikler systemer, databaser eller automatiserer prosesser som arbeider med designinformasjon og geometri.

Utvekslingsformatet som skal brukes både i beregningsavdelingene og på byggeplassen, bør ikke være basert på et spesifikt CAD- (BIM-) program. Geometrisk informasjon skal representeres i formatet direkte gjennom tessellering, uten referanse til den geometriske kjernen eller CAD-arkitektur.

Parametrisk geometri fra CAD kan betraktes som en mellomliggende kilde, men ikke som grunnlag for et universelt format. De fleste parametriske beskrivelser (inkludert BREP og NURBS) blir uansett konvertert til polygonal MESH for videre behandling. Hvis resultatet er det samme (tessellering og polygoner) og prosessen er enklere, er valget åpenbart. Dette er analogt med valget mellom grafontologier og strukturerte tabeller (som vi diskuterte i del fire): overdreven kompleksitet er sjeldent berettiget (fig. 3.2-10, fig. 6.1-8).

Åpne formater som OBJ, STL, glTF, SVF, CPIXML, USD og DAE, bruker en universell trekantnettstruktur, noe som gir dem betydelige fordeler. Disse formatene har utmerket interoperabilitet - de er enkle å lese og visualisere ved hjelp av tilgjengelige open source-biblioteker uten behov for komplekse, spesialiserte geometrikjerner som inneholder millioner av kodelinjer (figur 6.3-3). Disse allsidige geometriformatene brukes i alt fra relativt enkle kjøkkendesignverktøy i IKEA™ til komplekse objektvisualiseringssystemer i kino- og VR-applikasjoner. En viktig fordel er at det finnes et stort antall gratis biblioteker med åpen kildekode for å arbeide med disse formatene, og de er tilgjengelige for de fleste plattformer og programmeringsspråk.



Figur 6.3-3 Den samme geometrirepresentasjonen oppnås ved bruk av parametriske formater og geometrikjerner, eller ved bruk av triangulerte formater og visualiseringsbiblioteker med åpen kildekode.

I tillegg til brukerne selv har også CAD-leverandører problemer med å tolke utenlandske parametriske CAD-formater eller åpne IFC på grunn av ulike geometriske kjerner. I praksis bruker alle CAD-leverandører, uten unntakreverse engineering SDK til å overføre data mellom systemer, og ingen av dem er avhengige av formater som IFC eller USD [93] for interoperabilitetsformål.

I stedet for å bruke konsepter som fremmes av allianser av CAD-leverandører som de selv ikke bruker - er det mer produktivt for utviklere og brukere av CAD-løsninger å fokusere på å forstå fordelene ved hver tilnærming i en spesifikk sammenheng og å velge den ene eller andre typen geometri avhengig av brukstilfellet. Valget mellom ulike geometriske representasjoner er en avveining mellom nøyaktighet, beregningseffektivitet og de praktiske behovene i en bestemt oppgave.

Kompleksiteten knyttet til bruken av geometriske kjerner, som tradisjonelt har vært pålagt byggebransjen av store leverandører ved behandling av prosjekteringsdata, viser seg ofte å være overflødig. USD-formatet basert på MESH-geometri kan bli en slags "Pandoras eske" for bransjen, som åpner nye muligheter for utviklere til å organisere datautveksling - utenfor rammen av IFC og parametriske BREP-strukturer som er typiske for CAD-leverandører.

Når man ser nærmere på strukturen i USD, DAE, glTF, OBJ osv., blir det åpenbart at det finnes enklere, åpne formater som gjør det mulig å organisere overføring og bruk av geometrisk informasjon på en effektiv måte uten å måtte basere seg på kompleks parametrikk og lukkede geometriske kjerner. Denne tilnærmingen senker ikke bare den tekniske terskelen for utviklere, men fremmer også utviklingen av fleksible, skalerbare og reelt åpne løsninger for digital konstruksjon.

LOD, LOI, LOMD - unik klassifisering av detaljer i CAD (BIM)

I tillegg til geometriske representasjonsformater, og i en verden der ulike bransjer bruker ulike detaljeringsnivåer og datadybder, tilbyr CAD - (BIM-) metoder sine egne unike klassifiseringssystemer, som strukturerer tilnærmingen til å informere bygningsmodeller.

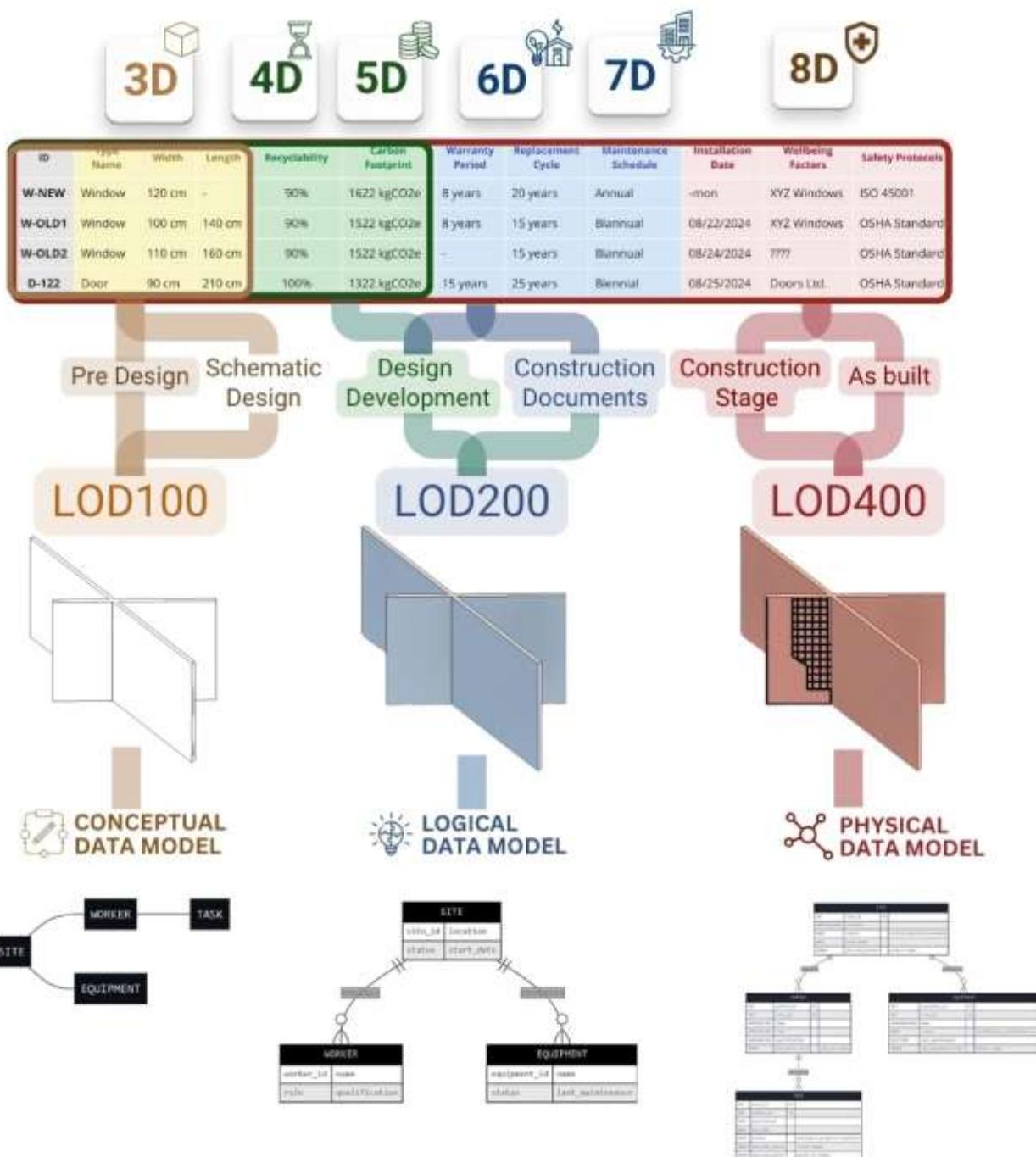
Et av eksemplene på nye tilnærninger til standardisering er innføringen av nivåer for modellutvikling, som gjenspeiler graden av beredskap og pålitelighet for både grafiske og informasjonsmessige komponenter. For å differensiere informasjonsinnholdet i arbeidet med CAD - (BIM-) data oppstod LOD (Level Of Detail) - detaljnivå for den grafiske delen av modellen, og LOI (Level Of Information) - nivå for datautvikling. I tillegg ble begrepet LOA (Level of Accuracy) introdusert for den integrerte tilnærmingen - nøyaktigheten til representerte elementer og LOG (Level of Geometry) for å bestemme nøyaktigheten til den grafiske representasjonen.

Detaljeringsnivåene (LOD) angis med tall fra 100 til 500, noe som gjenspeiler modellens utviklingsgrad. LOD 100 er en konseptuell modell med generelle former og dimensjoner. LOD 200 omfatter mer presise dimensjoner og former, men med betinget detaljering. LOD 300 er en detaljert modell med nøyaktige dimensjoner, former og elementplasseringer. LOD 400 inneholder detaljert informasjon som er nødvendig for fabrikasjon og installasjon av elementer. LOD 500 gjenspeiler anleggets faktiske tilstand etter bygging og brukes til drift og vedlikehold. Disse nivåene beskriver strukturen for informasjonsmetningen i en DAK-modell (BIM) på ulike stadier av livssyklusen, inkludert 3D, 4D, 5D og videre.

I virkelige prosjekter er det høye detaljnivået (LOD400) ofte for høyt, og det er tilstrekkelig å bruke LOD100-geometri eller til og med flate tegninger, mens resten av dataene kan innhentes enten beregningsmessig eller fra relaterte elementer som kanskje ikke har en distinkt geometri. For eksempel kan rom og romelementer (Room element categories) ikke ha noen visuell geometri, men inneholde betydelige mengder informasjon og databaser som mange forretningsprosesser er bygget opp rundt.

Det er derfor viktig å definere det nødvendige detaljnivået før man starter prosjekteringen. For 4D -7D er det ofte tilstrekkelig med DWG-tegninger og minimal LOD100-geometri. Den viktigste oppgaven i kravprosessen er å finne en balanse mellom modellens detaljrikdom og praktiske anvendelighet.

Hvis vi betrakter DAK-data (BIM) som en database (noe det er), er beskrivelsen av modellmetning gjennom nye akronymer i bunn og grunn ikke noe annet enn en trinnvis datamodellering for informasjonssystemer, fra det konseptuelle nivået til det fysiske (fig. 6.3-4), som ble diskutert i detalj i bokens tredje og fjerde del. Hver økning av LOD og LOI betyr at man legger til informasjon som er nødvendig for nye oppgaver: beregninger, byggeledelse, drift, og kjennetegnes ved at modellen suksessivt berikes med flere informasjonslag (3D -8D) i form av ulike parametere, som vi diskuterte i bokens femte del.



Figur 6.3-4 Prosessen med å konkretisere detaljene i et prosjekt er identisk med datamodellering fra en konseptuell til en fysisk datamodell.

Geometri er bare en del av designdataene, som det ikke alltid er behov for i byggeprosjekter, og hovedproblemet med å arbeide med CAD-data er ikke så mye hvordan modellene visualiseres, men mer hvordan dataene fra disse modellene kan brukes utenfor CAD- (BIM-) programmer.

På midten av 2000-tallet sto byggebransjen overfor en utfordring uten sidestykke med den raske økningen i datamengden i administrasjons- og databehandlingssystemer, spesielt de som kom fra CAD-avdelinger

(BIM). Denne dramatiske økningen i datamengden kom overraskende på bedriftslederne, og de var uforberedt på de økende kravene til datakvalitet og datahåndtering.

Nye CAD-standarder (BIM) - AIA, BEP, IDS, LOD, COBie

Ved å utnytte mangelen på åpen tilgang til CAD-databaser og den begrensede konkuransen i databehandlingsmarkedet, og ved hjelp av markedsføringskampanjene knyttet til det nye akronymet BIM, har organisasjoner som er involvert i utviklingen av metoder for håndtering av CAD-data, begynt å skape nye standarder og konsepter som de jure skal ha som mål å forbedre datahåndteringspraksis.

Selv om nesten alle initiativer som direkte eller indirekte støttes av CAD-leverandører og utviklere (BIM), har hatt som mål å optimalisere arbeidsflyten, har de resultert i en mengde standarder som ulike interessenter har kjempet for, noe som har ført til uklarhet og forvirring i byggebransjen når det gjelder dataprosesser.

La oss liste opp noen av de nye datastandardene, i tillegg til LOD, LOI, LOA, LOG, som har dukket opp i byggebransjen de siste årene:

- **BEP** (BIM Execution Plan) - beskriver hvordan CAD (BIM) skal integreres og brukes i et prosjekt, og definerer metoder og prosesser for datahåndtering.
- **EIR-dokument /AIA** (Opplysningskrav fra byggeren) - utarbeidet av byggeren før anbudskonkurransen og inneholder kravene til entreprenøren om å utarbeide og levere informasjon. Det fungerer som grunnlag for BEP i det respektive prosjektet.
- **AIM** (Asset Information Model) er en del av BIM-prosessen. Når prosjektet er levert og ferdigstilt, kalles datamodellen for Asset Information Model eller AIM. Formålet med AIM er å forvalte, vedlikeholde og driftet den realiserte eiendelen.
- **IDS** (Information Delivery Specification) - definerer kravene til og hvilke data og i hvilket format som kreves i ulike faser av et byggeprosjekt.
- **iLOD** er detaljnivået LOD, som informasjonen er representert med i BIM-modellen. Den definerer hvor detaljert og fullstendig informasjonen i modellen er, fra grunnleggende geometriske representasjoner til detaljerte spesifikasjoner og data.
- **eLOD** - LOD detaljnivå for de enkelte elementene i en DAK-modell (BIM). Det definerer i hvilken grad hvert element er modellert og den tilhørende informasjonen, for eksempel dimensjoner, materialer, ytelsesegenskaper og andre relevante attributter.
- **APS** (Platform Services) og andre produkter fra store CAD-leverandører (BIM) - beskriver verktøyene og infrastrukturen som trengs for å lage koblede og åpne datamodeller.

Selv om det uttalte formålet med å implementere CAD (BIM)-standarder - som LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - er å forbedre kvaliteten på datahåndteringen og utvide automatiseringsmulighetene, fører bruken av dem i praksis ofte til overdreven kompleksitet og fragmentering av prosesser. Hvis vi betrakter en CAD-modell (BIM) som en slags database, blir det åpenbart at mange av disse standardene dupliserer veletablerte og effektive tilnærminger som brukes i andre bransjer i arbeidet med informasjonssystemer. I stedet for å forenkle og forene, skaper slike initiativer ofte

en ekstra terminologisk byrde og hindrer implementeringen av virkelig åpne og fleksible løsninger.

Mange av disse nye konseptene erstatter faktisk modellerings- og datavalideringsprosessene som ble diskutert i detalj i de første delene av boken, og som lenge har vært brukt i andre sektorer av økonomien. I bygg- og anleggsbransjen går standardiseringsprosessen derimot ofte i motsatt retning - det skapes nye formater for databeskrivelser, nye standarder og nye konsepter for datavalidering, som ikke alltid fører til reell ensartethet og praktisk anvendelighet. I stedet for å forenkle og automatisere saksbehandlingen, blir bransjen derfor utsatt for flere nivåer av regulering og byråkrati (fig. 6.3-1), noe som ikke alltid bidrar til økt effektivitet.

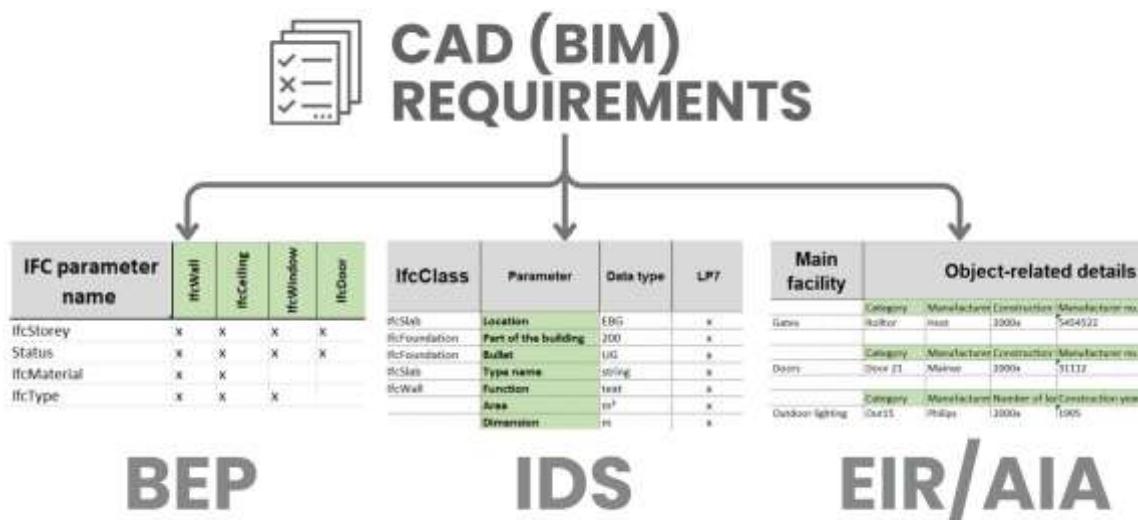


Fig. 6.3-1 Kravene til data- og informasjonsinnhold er redusert til beskrivelsen av attributter og deres grenseverdier, beskrevet ved hjelp av tabeller.

I stedet for å forenkle databehandlingen, fører nye konsepter knyttet til CAD-data (BIM) ofte til ytterligere kompleksitet og tvister allerede på tolkningsstadiet og i forbindelse med grunnleggende definisjoner.

Et av de nyeste eksemplene på nye konsept er IDS-formatet (introdusert i 2020) som gjør det mulig å beskrive kravene for attributsammensetningen til en informasjonsmodell i åpen BIM-konseptet. IDS-krav beskriver informasjon om attributter og deres grenseverdier i form av en strukturert tabell (Excel eller MySQL), som deretter oversettes til markeringen av et semi-strukturert XML-format, omdøpt fra XML til den spesielle forkortelsen IDS.

I motsetning til det synet som fremmes av leverandører og støttes av BIM og åpen BIM, om at datahåndtering i byggebransjen er unik på grunn av bruken av spesialiserte verktøy som CAD og BIM, er ikke dataformatene og datahåndteringspraksisen i denne bransjen annerledes enn andre bransjer.

Antallet krav til prosjekter og CAD (BIM) kan forenkles ved å bruke en enkelt kravtabell med attributtkolonner, som beskrevet i kapittelet "Oversette krav til strukturert form", uten å måtte oversette de opprinnelig strukturerde kravene til formater som ikke er tabellformater (IDS er i utgangspunktet beskrevet via en tabell).

Den forenklede tilnærmingen (figur 6.3-2), som inkluderer kolonner for entitetsidentifikatorer, egenskaper og grenseverdier som ble diskutert i detalj i tidligere kapitler (figur 4.4-9, figur 4.4-16, figur 7.3-10), eliminerer behovet for å konvertere krav til IDS-XML-format. Denne metoden gir en direkte, mindre tungvint og mer transparent mekanisme for datakvalitetskontroll. Den baserer seg på mye brukte verktøy, fra regulære uttrykk (RegEx) til dataframes, Pandas og standard ETL -payplanes - akkurat som de som brukes av fagfolk i andre sektorer av økonomien til å arbeide med data.

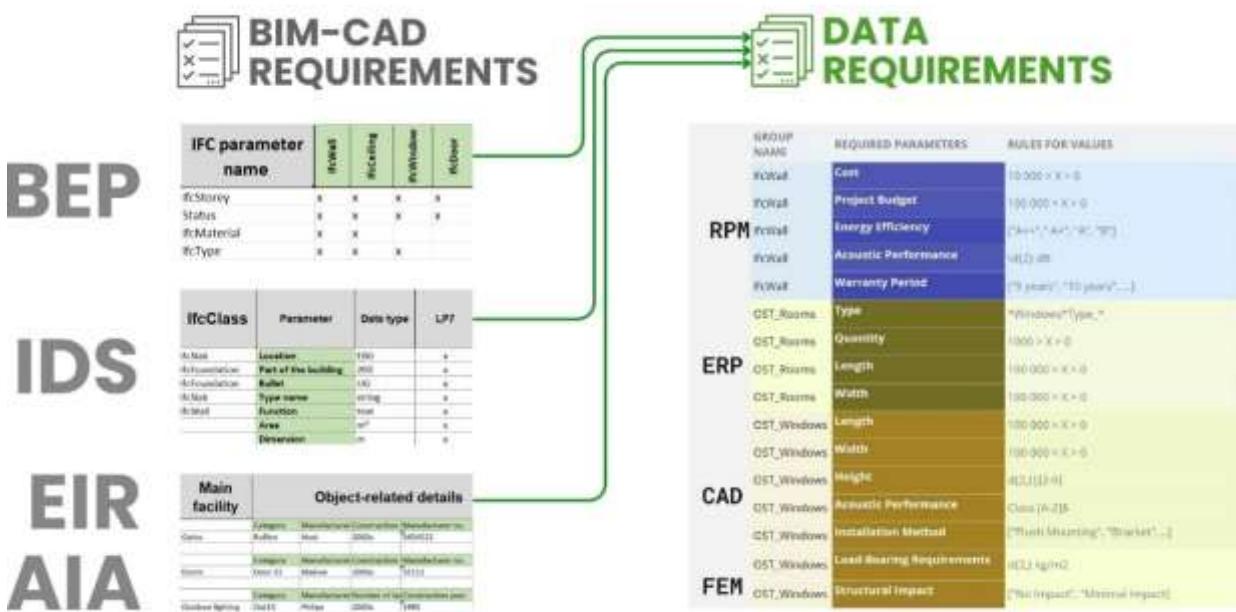


Fig. 6.3-2 Datakrav i andre bransjer er forenklet til en strukturert beskrivelse av attributter og deres grenseverdier.

På grunn av dataenes lukkede natur har det over tid dukket opp stadig flere nye tilnæringer og teknikker for å kontrollere og håndtere disse mangfoldige dataene i byggebransjen, selv om dataene i byggeprosjekter i hovedsak er de samme som i andre bransjer. Mens andre bransjer har klart seg med standardiserte tilnærninger til databehandling, fortsetter bygg- og anleggsbransjen å utvikle nye og unike dataformater, krav og valideringskonsepter.

Metodene og verktøyene som brukes til å samle inn, bearbeide og analysere data i bygg- og anleggsbransjen, bør ikke være fundamentalt forskjellige fra dem som brukes av spesialister i andre sektorer av økonomien.

Bransjen har utviklet et eget terminologisk økosystem som krever kritisk refleksjon og revurdering.

- STEP-formatet er plassert under det nye navnet IFC, supplert med konstruksjonskategorisering, uten å ta hensyn til begrensningene i selve STEP-formatet.

- Det parametriske formatet IFC brukes i datakommunikasjonsprosesser til tross for at det mangler en enhetlig geometrisk kjerne som er nødvendig for visualisering og beregning.
- Tilgang til databaser CAD -systemer markedsføres under betegnelsen "BIM ", uten at det gis noen nærmere beskrivelse av disse databasene og tilgangen til dem.
- Leverandørene fremmer interoperabilitet gjennom formatene IFC og USD, ofte uten å omsette dem i praksis, ved hjelp av kostbar reverse engineering som de selv har slitt med.
- Begrepene LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD brukes universelt for å beskrive de samme enhetsparametrene, uten referanse til modellerings- og verifiseringsverktøy som lenge har vært brukt i andre bransjer.

Byggebransjen viser at selv om det høres rart ut, er alt det ovennevnte mulig i byggebransjen - spesielt hvis hovedmålet er å tjene penger på hvert trinn i databehandlingen gjennom salg av spesialiserte tjenester og programvare. Fra et forretningsmessig synspunkt er det ikke noe galt med dette. Men det er fortsatt et åpent spørsmål om slike akronymer og tilnærminger knyttet til DAK (BIM) virkelig tilfører verdi og forenkler de profesjonelle prosessene.

I byggebransjen fungerer et slikt system fordi bransjen selv tjener det meste av sin spekulative profitt i denne labyrinten av systemer og akronymer. Selskaper som er interessert i transparente prosesser og åpne data, er sjeldne. Denne komplekse situasjonen vil trolig fortsette i det uendelige - helt til kunder, klienter, investorer, banker og private equity-selskaper begynner å kreve en tydeligere og mer informert tilnærming til informasjonshåndtering.

Bransjen har samlet på seg et stort antall akronymer, men de beskriver alle de samme prosessene og datakravene i varierende grad. Det er tvilsomt om de virkelig bidrar til å forenkle arbeidsflyten.

Mens konsepter og markedsføringsakronymer kommer og går, vil selve valideringsprosessene for datakrav for alltid forbli en integrert del av forretningsprosessene. I stedet for å opprette flere og flere spesialiserte formater og forskrifter på, bør byggebransjen se til verktøy som allerede har vist seg å være effektive på andre områder som finans, industri og IT.

Overfloden av begreper, akronymer og formater skaper en illusjon av gjennomarbeidede digitale byggeprosesser. Men bak markedsføringskonsepter og kompleks terminologi skjuler det seg ofte en enkel, men ubeleilig sannhet: Data er fortsatt vanskelig tilgjengelige, dårlig dokumentert og strengt knyttet til spesifikke programvareløsninger.

For å komme ut av denne onde sirkelen av akronymer og formater for formatenes skyld, er det nødvendig å se på CAD (BIM) -systemer ikke som magiske verktøy for informasjonshåndtering, men som det de egentlig er - spesialiserte databaser. Og det er gjennom dette prismet at man kan forstå hvor markedsføringen slutter og det virkelige arbeidet med informasjon begynner.



KAPITTEL 6.4.

DESIGNPARAMETRISERING OG BRUK AV LLM FOR CAD-DRIFT

Illusjonen om at CAD-data er unike (BIM): veien til analyse og åpne formater

Moderne CAD-plattformer (BIM) har i betydelig grad endret tilnærmingen til design og informasjonshåndtering i byggeprosjekter. Mens disse verktøyene tidligere hovedsakelig ble brukt til å lage tegninger og 3D-modeller, fungerer de i dag som fullverdige prosjektdatalagre. I henhold til Single Source of Truth-konseptet blir den parametriske modellen i stadig større grad den viktigste og ofte den eneste kilden til prosjektinformasjon, noe som sikrer dens integritet og relevans gjennom hele prosjektets livssyklus.

Den viktigste forskjellen mellom CAD - (BIM -) plattformer og andre systemer for håndtering av byggdata er behovet for spesialiserte verktøy og API-er for å få tilgang til informasjonen (den eneste sannhetskilden). Disse databasene er ikke universelle i tradisjonell forstand: I stedet for en åpen struktur og fleksibel integrasjon er de et lukket miljø, knyttet til en bestemt plattform og et bestemt format.

Til tross for kompleksiteten i arbeidet med CAD -data er det et viktigere spørsmål som går ut over den tekniske gjennomføringen: Hva er egentlig CAD-databaser (BIM)? For å svare på dette spørsmålet er det nødvendig å gå utover de vanlige akronymene og konseptene som programvareutviklere pålegger oss. I stedet er det verdt å fokusere på essensen i arbeidet med prosjektinformasjon: data og behandlingen av dem.

Forretningsprosessen i bygg- og anleggsbransjen starter ikke med arbeid i CAD - eller BIM - verktøy, men med utforming av prosjektkrav og datamodellering. Først defineres oppgaveparametrene: listen over enheter, deres opprinnelige egenskaper og grenseverdier som må tas i betraktning når en bestemt oppgave skal løses. Først da opprettes modeller og elementer i CAD (BIM)-systemer på grunnlag av de spesifiserte parameterne

Prosesssen som går forut for opprettelsen av informasjon i CAD - (BIM-) databaser er helt den samme som datamodellingsprosessen som ble diskutert i detalj i bokens fjerde del og kapittelet "Datamodellering: konseptuell, logisk og fysisk modell" (fig. 4.3-1).

På samme måte som vi i datamodellering lager krav for dataene som vi senere ønsker å behandle i databasen, lager CAD-databaseledere designkrav i form av flere tabellkolonner eller lister over nøkkelverdipar (fig. 6.4-1, trinn 1-2). Og bare på grunnlag av disse innledende parametrene ved hjelp av API automatisk eller manuelt, oppretter designeren (eller rettere sagt foredler) objekter i CAD- (BIM) databaser (trinn 3-4), hvoretter de blir sjekket igjen for å overholde de opprinnelige kravene (trinn 5-6). Denne prosessen - definisjon→ oppretting→ validering→ justering (trinn 2-6) - gjentas iterativt til datakvaliteten, akkurat ved datamodellering, når det ønskede nivået for målsystemet - dokumenter, tabeller eller dashbord (trinn 7).

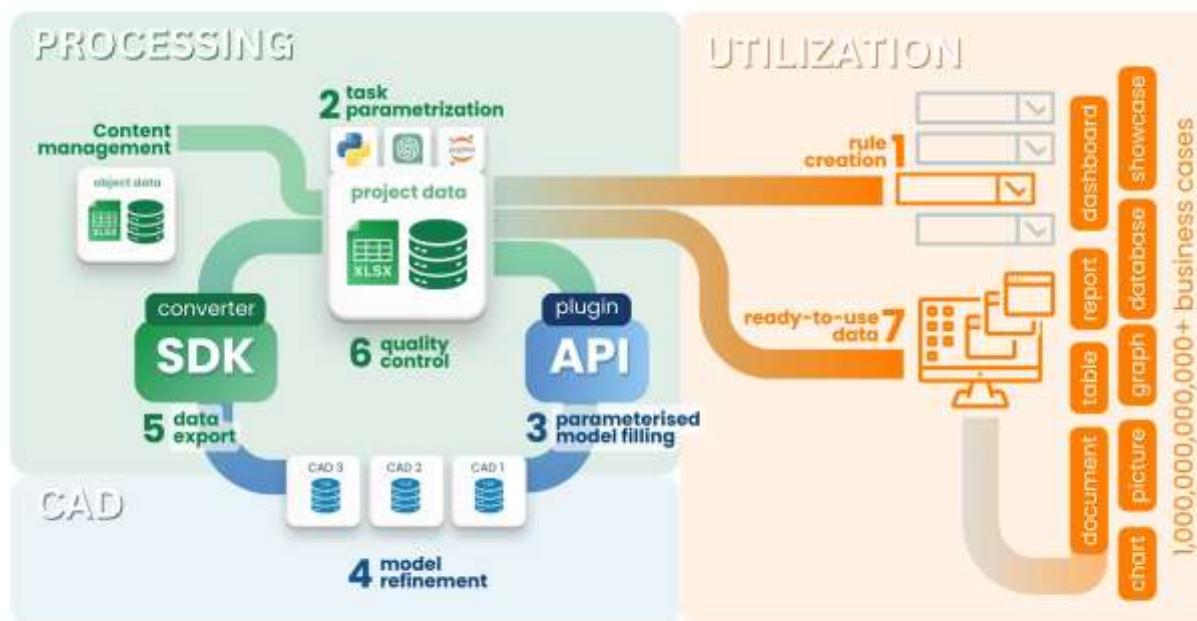


Fig. 6.4-1 Syklusen for informasjonsmetning av databaser for forretningsprosesser i gjennomføringen av byggeprosjekter.

Hvis vi betrakter CAD (BIM) som en mekanisme for parameteroverføring i form av et sett med nøkkelverdipar generert på grunnlag av krav som er definert utenfor prosjekteringsmiljøet (fig. 6.4-1, trinn 1-2), flyttes fokus i diskusjonen fra spesifikke programvarelösninger og deres begrensninger til mer grunnleggende aspekter - datastruktur, datamodeller og datakrav. I bunn og grunn snakker vi om parametermetning av databasen og den klassiske datamodelleringsprosessen (trinn 2-3 og 5-6). Den eneste forskjellen er at på grunn av CAD-databasenes lukkede natur og de særegne formatene som brukes, ledsages denne prosessen av bruk av spesialiserte BIM-verktøy. Spørsmålet er hva som er det unike med BIM, hvis det ikke finnes lignende tilnæringer i andre branjer?

I løpet av de siste 20 årene har BIM blitt posisjonert som mer enn bare en enkelt datakilde. CAD-BIM-pakken markedsføres ofte som et parametrisk verktøy med en iboende integrert database [64], som er i stand til å automatisere prosessene for design, modellering og livssyklusadministrasjon av byggeprosjekter. I realiteten har BIM imidlertid blitt mer et verktøy for å holde brukerne på leverandørens plattform enn en praktisk metode for data- og prosesshåndtering.

Resultatet er at CAD- (BIM-) data er isolert på plattformene deres, og prosjektinformasjon skjules bak proprietære API-er og geometrikjerner. Dette har fratatt brukerne muligheten til å få uavhengig tilgang til databaser og trekke ut, analysere, automatisere og overføre data til andre systemer, utenom leverandørenes økosystemer.

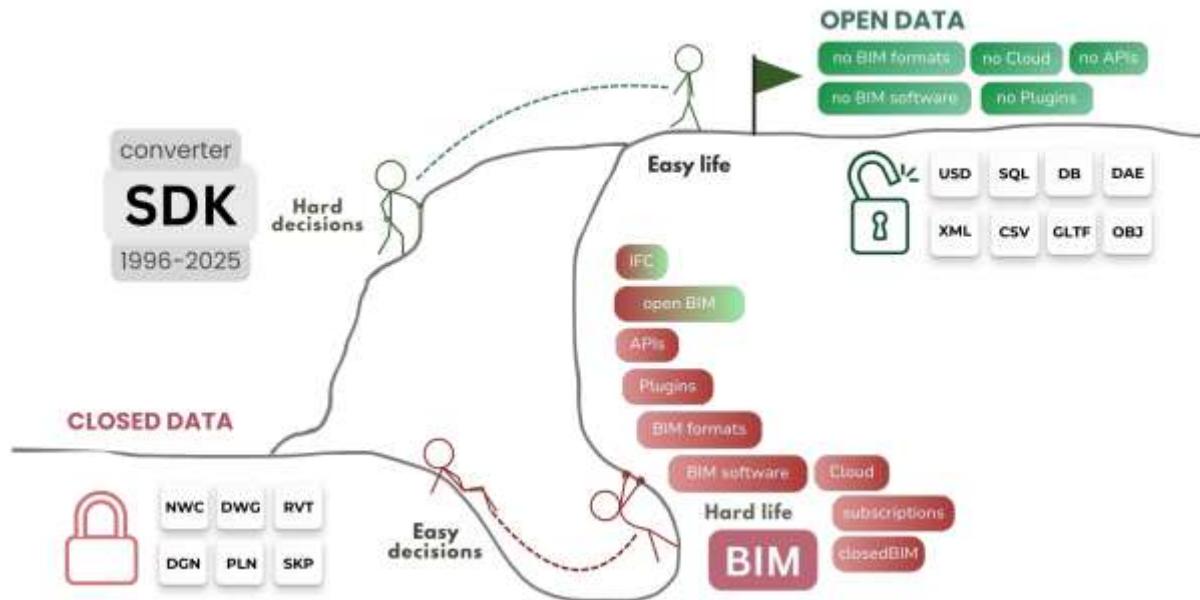


Fig. 6.4-2 I bygg- og anleggsbransjen krever moderne formater sofistikerte geometrikjerner, et årlig oppdatert API og spesielle lisenser for CAD -(BIM-) programvare.

Bedrifter som arbeider med moderne DAK-verktøy, bør bruke den samme tilnærmingen til arbeid med data som alle DAK-leverandører uten unntak selv bruker i praksis: datatransformasjon ved hjelp av SDK - reverse engineering-verktøy, som DAK-leverandørene har kjempet mot siden 1995 [75]. Med full tilgang til CAD-databasen og ved hjelp av reverse engineering-verktøy kan vi få [127] et flatt sett med entiteter med attributter og eksportere dem til et hvilket som helst praktisk åpent format (figur 6.4-2), inkludert både geometri og parametere for designelementer. Denne tilnærmingen endrer paradigmet for arbeid med informasjon fundamentalt - fra filorientert til datasentrisk arkitektur:

- Dataformater som RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET og andre inneholder identisk informasjon om elementer i det samme prosjektet. Dette betyr at kunnskap om et bestemt format og dets skjema ikke bør være en hindring for å arbeide med selve dataene.
- Data fra alle formater kan kombineres til én enkelt åpen strukturert og granulær struktur (figur 9.1-10) som inneholder MESHs trekantede geometri og egenskapene til alle objektenheter, uten begrensningene fra geometriske kjerner.
- Dataanalyse etterstreber universalitet: Ved hjelp av åpne data kan du arbeide med prosjektdata uavhengig av hvilket format som brukes.
- Minimering og avhengighet av API-er og plugins fra leverandører: Arbeid med data er ikke lenger avhengig av API-kompetanse.

Når og CAD-datakrav omdannes til strukturerte representasjonsformater som er enkle å analysere, er utviklerne ikke lenger avhengige av spesifikke dataskjemaer og lukkede økosystemer.

Design gjennom parametere: fremtiden for CAD og BIM

Ingen byggeprosjekter i verden har noen gang startet i et CAD-program. Før en tegning eller modell tar form i DAK, går den gjennom konseptualiseringfasen (fig. 6.4-1, trinn 1-2), der fokuset ligger på de parameterne som definerer grunnideen og logikken i det fremtidige objektet. Dette stadiet tilsvarer det konseptuelle nivået i datamodellering (fig. 4.3-6). Parametrene kan eksistere utelukkende i designerens hode, men ideelt sett er de ordnet i form av strukturerte lister, tabeller eller lagret i databaser (fig. 6.4-3), noe som muliggjør transparens, reproducertbarhet og ytterligere automatisering av designprosessen.

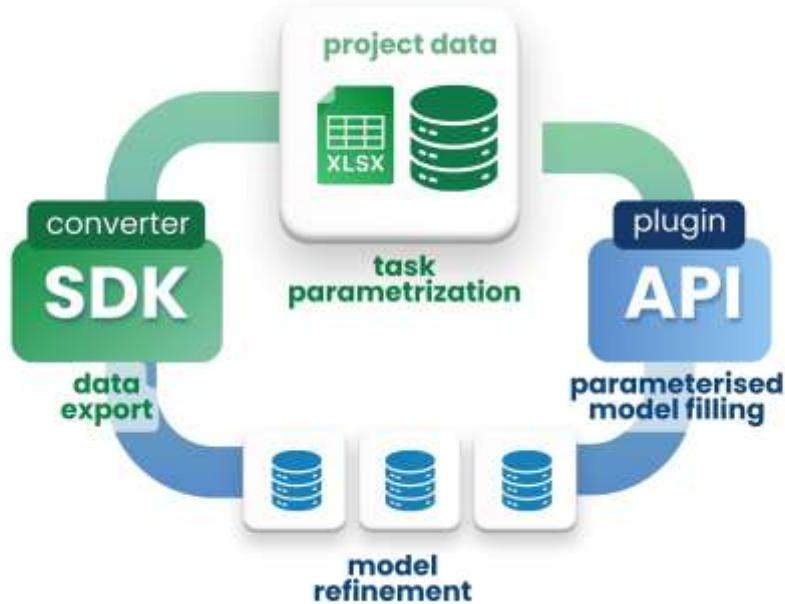


Fig. 6.4-3 Designprosessen er en iterativ prosess der CAD-databasen fylles opp med informasjon utenfra ved hjelp av krav i verdikjeden.

Før selve CAD-modelleringen (den logiske og fysiske fasen av datamodelleringen (fig. 4.3-7)) starter, er det viktig å definere de grenseparametrene som danner grunnlaget for prosjektet. Disse attributtene, i likhet med andre krav, samles inn helt fra slutten av dataanvendelseskjeden (f.eks. systemer), og gjennom dem er begrensningene, målene og nøkkelegenskapene til fremtidige objekter i prosjektet allerede definert.

Selve modelleringen kan automatiseres med 60-100% ved hjelp av parametriske modellingsverktøy (fig. 6.4-3), forutsatt at kravene er veldefinerte. Så snart prosjektet er beskrevet i form av parametere, blir det teknisk mulig å utforme det, for eksempel ved hjelp av visuelle programmeringsspråk som Grasshopper Dynamo, innebygd i moderne CAD -miljøer eller gratisløsninger i Blender, UE, Omniverse.

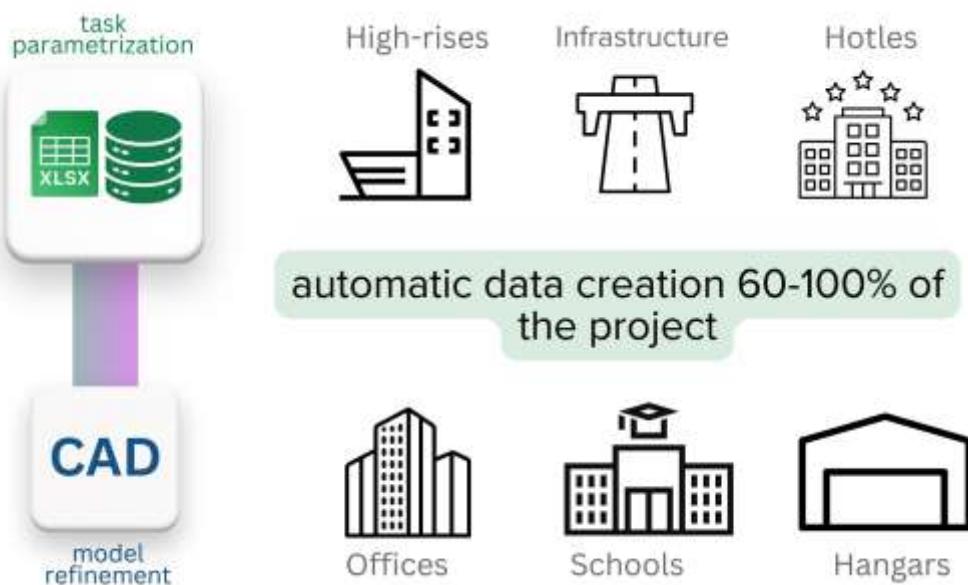


Fig. 6.4-4 De fleste typifiserte prosjekter opprettes allerede i dag helt automatisk takket være parametriske programmeringsverktøy.

Allerede i dag skapes store industrielle og typifiserte prosjekter ikke av designavdelingens hender, men ved hjelp av parametriske verktøy og visuell programmering. Dette gjør det mulig å bygge en modell basert på data i stedet for på subjektive beslutninger fra en bestemt designer eller leder.

Innhold går forut for design. Design uten innhold er ikke design, men dekorasjon [128].

- Jeffrey Zeldman, webdesigner og gründer

Prosessen starter ikke med tegning eller 3D -modellering, men med utforming av krav. Det er kravene som bestemmer hvilke elementer som skal brukes i prosjektet, hvilke data som må overføres til andre avdelinger og systemer. Bare eksistensen av strukturerte krav gjør det mulig å automatisk kontrollere modeller med jevne mellomrom (for eksempel hvert 10. minutt uten å distrahere designeren fra arbeidet).

Kanskje vil CAD- (BIM-) fremtiden bare bli et grensesnitt for å fylle databasen, og hvilket CAD-verktøy modelleringen utført i (fysisk nivå) vil ikke ha noen betydning.

På samme måte brukes 3D-modellering ofte innen maskinteknikk, men det er ikke et nødvendig eller obligatorisk element i prosjektet. I de fleste tilfeller er den klassiske 2D-dokumentasjonen tilstrekkelig, og den nødvendige informasjonsmodellen opprettes på grunnlag av denne. Denne modellen settes sammen av komponenter som er strukturert i henhold til industristandarder, og inneholder all nødvendig informasjon for å forstå design- og produksjonsorganisasjonen. Fabrikkinformasjonsmodellen brukes deretter til å lage en fabrikkinformasjonsmodell, der spesifikke produkter og flytdiagrammer, som allerede er tilpasset teknologenes behov, legges til. Hele prosessen kan organiseres uten unødvendig kompleksitet, og uten å overbelaste systemet med 3D-grafikk som ikke gir reelle fordeler.

Det er viktig å forstå at selve 3D-modellen og CAD-systemet ikke skal spille hovedrollen - det er bare et verktøy for kvantitativ og geometrisk analyse. Alle andre parametere, bortsett fra geometri, som beskriver enheten, bør lagres og behandles utenfor CAD-miljøet hvis det er mulig (BIM).

Design gjennom parametere er ikke bare en trend, men den unngåelige fremtiden for byggebransjen. I stedet for å lage komplekse 3D-modeller manuelt, vil designere jobbe med data, validere dem og automatisere prosessene, noe som vil bringe bygg- og anleggsbransjen nærmere programmeringsverdenen. Over tid vil designprosessene bli bygget på prinsippene for programvareutvikling:

- Opprett krav → Opprett modell → Last opp til server → Valider endringer → Pull-forespørsel
- Pull-forespørsler kjører automatisk modellkontroller mot krav som ble opprettet før eller under utformingen som en del av Pull-forespørselen
- Etter datakvalitetskontroller og godkjenning, blir endringene implementert i prosjektet, den felles databasen eller overført automatisk til andre systemer

Allerede i dag starter slike designendringer i maskinindustrien med en endringsmelding. En lignende ordning venter byggebransjen: Prosjekteringen vil bli en iterativ prosess der hvert trinn støttes av parametriske krav. Et slikt system vil gjøre det mulig for designere å opprette automatiserte kontroller og automatiserte pull-forespørsler for spesifikke krav.

Fremtidens designer er først og fremst en dataoperatør, ikke en manuell modellerer. Hans oppgave er å fylle prosjektet med parametriske enheter, der geometri bare er ett av attributtene.

Det er forståelsen for betydningen av datamodellering, klassifisering og standardisering, som har blitt diskutert i detalj i de foregående kapitlene i boken, som vil spille en viktig rolle i transformasjonen. Fremtidens designforskrifter vil bli formalisert som nøkkel-verdi-parameterpar i form av XLSX eller XML-skjemaer.

Fremtidens byggebransje handler om å samle inn data, analysere dem, validere dem og automatisere prosesser ved hjelp av analyseverktøy. BIM (eller CAD) er ikke det endelige målet, men bare et stadium i utviklingen. Når fagfolk innser at de kan arbeide direkte med data, uten bruk av tradisjonelle CAD-verktøy, vil selve begrepet "BIM" gradvis vike for konseptene med bruk av strukturerte og detaljerte data om byggeprosjekter.

En av de viktigste faktorene som har fremskyndet denne utviklingen, er fremveksten av store språkmodeller (LLM) og verktøyene som er basert på disse. Disse teknologiene endrer måten designdata håndteres på, og gir tilgang til informasjon uten behov for inngående kunnskap om API-er eller leverandør løsninger. Med LLM-er blir prosessen med å opprette et krav og samhandle med CAD data intuitiv og tilgjengelig.

Fremveksten av LLM i design CAD databehandlingsprosesser

I tillegg til utviklingen av verktøy for tilgang til CAD-databaser og åpne og forenklede CAD -formater, har fremveksten av LLM -verktøy (Large Language Models) revolusjonert behandlingen av designdata. Mens tilgangen til informasjon tidligere hovedsakelig gikk gjennom komplekse grensesnitt og krevde programmeringsferdigheter og kunnskap om API, er det nå mulig å samhandle med data ved hjelp av naturlig språk.

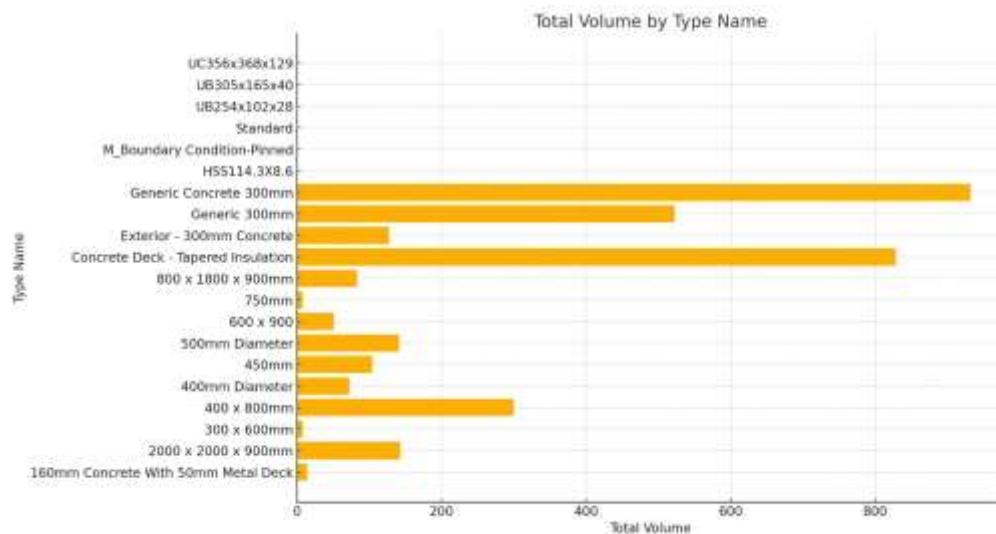
Ingenører, ledere og planleggere uten teknisk bakgrunn kan få nødvendig informasjon fra prosjektdaten ved å formulere spørsmål på vanlig språk. Forutsatt at dataene er strukturerte og tilgjengelige (fig. 4.1-13), er det nok å stille et spørsmål i LLM chat, for eksempel: "Vis i en tabell med gruppering etter type alle vegg med et volum på mer enn 10 kubikkmeter" - og modellen vil automatisk konvertere denne spørringen til SQL eller kode i Pandas, og generere en oppsummeringstabell, en graf eller til og med et ferdig dokument.

Nedenfor følger noen eksempler fra virkeligheten på hvordan LLM -modeller samhandler med designdata som er representert i ulike CAD- (BIM-) formater.

- Eksempel på en spørring i LLM chat til et DAK-prosjekt i RVT -format etter konvertering (fig. 4.1-13) til en tabellarisk dataramme (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller en hvilken som helst annen):

Grupper dataene i datarammen hentet fra RVT-filen etter "Type name" når du summerer parameteren "Volume" og vis antall elementer i gruppen. Og vis alt dette som et horisontalt histogram uten nullverdier.

- LLM-respons som et horisontalt søylediagram (PNG-format):



Figur 6.4-5 I stedet for 17 museklikk eller 40 linjer med kode ved bruk av plugin-moduler, kan vi i LLM hente QTO-tabellen umiddelbart ved hjelp av en tekstforespørsel.

- For å generere en QTO -tabell med veggtyper med totalt areal og antall fra kategorien "Vegger", la oss formulere en tekstforespørsel til LLM -chat:

Ta kun de elementene fra prosjektdatarammen som har "OST_Walls" i "Category"-parameteren, gruppér dem etter "Type Name", summer verdien i "Area"-kolonnen, legg til antallet og vis dem i tabellen ved å fjerne nullverdier.

- 💡 LLM-svar i form av en ferdig QTO-tabell:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

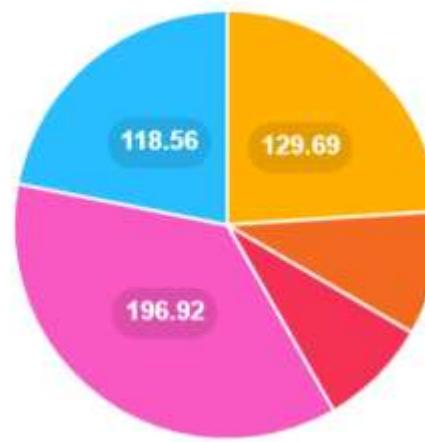
Fig. 6.4-6 Opprettelse av en QTO tabell i naturlig språk gir samme kvalitetsresultat som ved bruk av CAD - (BIM-) verktøy.

- 💡 La oss spørre prosjektet i IFC -format etter å ha konvertert det til en tabellarisk dataramme og legge inn en lignende tekstforespørsel i en hvilken som helst LLM chat:

Ta bare de elementene fra prosjektet som har nivå 1- og nivå 2-verdier i parameteren "Parent", og ta elementene som har IfcSlab-verdier i parameteren "Category", gruppér deretter disse elementene etter parameteren "ObjectType", summer verdiene i parameteren "PSet_RVT _Dimensions Area", og vis dem som et kakediagram.

■ LLM-respons som et ferdig kakediagram over elementgrupper fra IFC-data :

For ■ Floor:127mm Slab on Grade, ■ Floor:150mm Exterior Slab on Grade,
 ■ FloorFinish Floor - Ceramic Tile, ■ Floor:Finish Floor - Wood, and
 ■ Floor:Residential - Wood Joist with Subflooring



Figur 6.4-7 Resultatet av en spørring etter IFC -data i et strukturert format kan være en hvilken som helst type graf som er praktisk for å forstå dataene.

Bak hver av de ferdige løsningene (fig. 6.4-5 - fig. 6.4-7) ligger det et dusin linjer med Python-kode som bruker Pandas-biblioteket. Den resulterende koden kan kopieres fra LLM-chatrommet og brukes i en hvilken som helst lokal eller nettbasert IDE for å få identiske resultater utenfor LLM-chatrommet.

I samme LLM-chat kan vi ikke bare arbeide med prosjekter hentet fra 3D-CAD (BIM) formater, men også med flate tegninger i DWG-format, som vi kan spørre LLM chat om å vise for eksempel data om elementgrupper i form av linjer eller 3D-geometrier etter konvertering til en strukturert form.

Automatisert analyse av DWG -filer med LLM og Pandas

Prosessen med å behandle data fra DWG -filer har alltid vært en kompleks oppgave som krever spesialisert programvare og ofte manuell analyse, på grunn av informasjonens ustrukturerte natur. Med utviklingen av kunstig intelligens og LLM-verktøy har det imidlertid blitt mulig å automatisere mange av trinnene i denne prosessen, som i dag for det meste er manuell. La oss vurdere en ekte Pipeline av forespørsler til LLM (i dette eksemplet ChatGPT) for å jobbe med DWG-tegninger, som lar deg jobbe med prosjektet:

- Filtrer DWG-data etter lag, ID og koordinater
- Visualiser geometrien til elementene
- Automatisk kommentering av tegninger basert på parametere
- Utvid veggpolylinjene til horisontalplanet
- Lag interaktive 3D -visualiseringer av plane data

■ Strukturer og analyser konstruksjonsdata uten komplekse CAD-verktøy

I vårt tilfelle starter prosessen med å bygge Pipelinesekvensiell kodegenerering via LLM. Først genereres en spørring som beskriver oppgaven. ChatGPT genererer Python -kode, som kjøres og analyseres, og viser resultatet inne i chatterommet. Hvis resultatet ikke er som forventet, korrigeres forespørselen, og prosessen gjentas

Pipeline er en sekvens av automatiserte trinn som utføres for å behandle og analysere data. I en slik prosess tar hvert trinn data som input, utfører transformasjoner og sender resultatet videre til neste trinn.

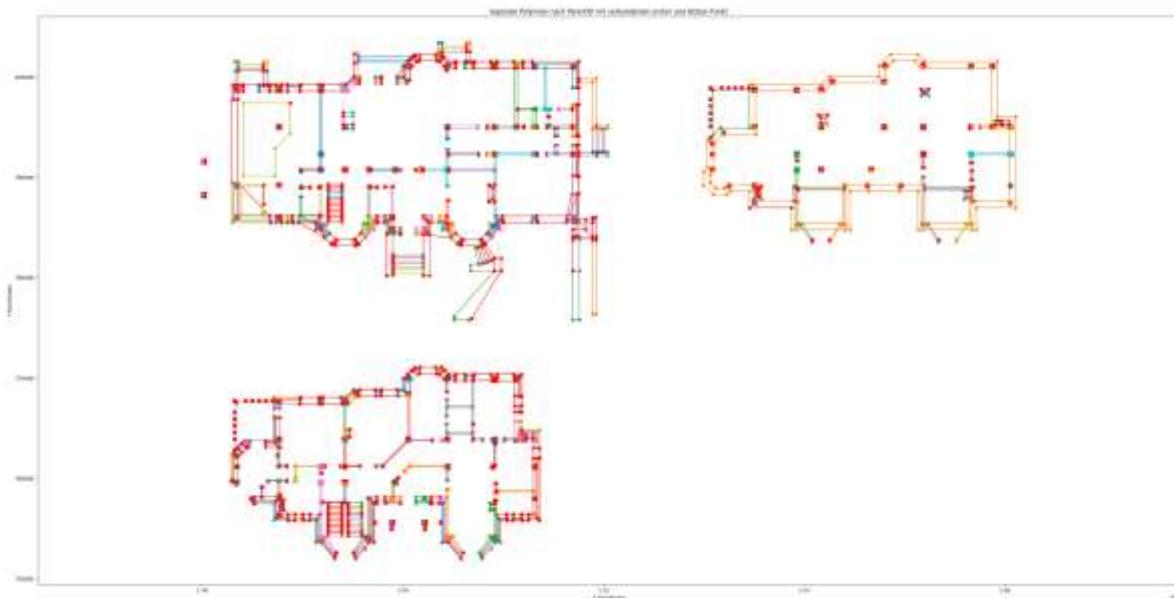
Etter å ha oppnådd ønsket resultat kopieres koden fra LLM og limes inn i koden i form av blokker i en av de praktiske IDE-ene, i vårt tilfelle på Kaggle-plattformen.com. De resulterende kodelinjene kombineres til én enkelt pipeline, som automatiserer hele prosessen - fra datainnlasting til den endelige analysen. Denne tilnærmingen gjør det mulig å utvikle og skalere analytiske prosesser raskt, uten dyp programmeringskompetanse. Den fullstendige koden til alle fragmentene nedenfor, sammen med eksempler på spørninger, finner du på Kaggle.com-plattformen ved å søke på "DWG Analyse with ChatGPT | DataDrivenConstruction" [129].

La oss starte prosessen med å jobbe med DWG-data, etter konvertering til strukturert form (Fig. 4.1-13), med et klassisk trinn - gruppering og filtrering av alle tegningsdata, nødvendig for vår oppgave veggelementer, spesielt polylinjer (parameter 'ParentID' gjør det mulig å gruppere linjer i grupper), som i parameteren (dataframe kolonne) "Lag" har en strengverdi som inneholder følgende bokstavkombinasjon (RegEx) - "vegg".

- ❷ For å få koden for en lignende oppgave og resultatet i form av et bilde, bør du skrive følgende spørring i LLM:

Først sjekker du om datarammen som er hentet fra DWG, inneholder de definerte kolonnene: "Layer", "ID", "ParentID" og "Point". Filtrer deretter ut ID-ene fra kolonnen "Layer" som inneholder strengen "wall". Finn elementene i kolonnen "ParentID" som samsvarer med disse identifikatorene. Definer en funksjon for å rense og splitte dataene i kolonnen "Point". Dette inkluderer å fjerne parenteser og dele opp verdiene i x-, y- og z-koordinater. Plott dataene ved hjelp av matplotlib. Tegn en egen polylinje som forbinder "Point"-koordinatene for hver unike "ParentID". Sørg for at det første og siste punktet er koblet sammen hvis mulig. Angi passende etiketter og titler, og sørge for at x- og y-aksene er skalert likt.

- ❸ Svaret LLM gir deg et ferdig bilde, og bak dette ligger Python-koden som genererte det:



Figur 6.4-8 LLM koden hentet ut alle linjene i "wall"-laget fra DWG-filen, renset koordinatene og konstruerte polylinjene ved hjelp av et av Python-bibliotekene.

- La oss nå legge til arealparameteren som hver polylinje har i sine egenskaper (i en av datarammekolonnene):

Finn nå bare én "ParentID" fra hver polylinje - finn denne ID-en i "ID"-kolonnen, ta "Area"-verdien, del på 1 000, og legg denne verdien til i grafen

- LLM-svaret viser en ny graf der hver polylinje har en bildetekst med arealet:



Fig. 6.4-9 LLM har lagt til kode som tar arealverdiene for hver polylinje og legger dem til i bildet med linjevisualisering.

- Deretter konverterer vi hver polylinje til en horisontal linje, legger til en parallel linje i en høyde på 3000 mm og kobler dem sammen til et enkelt plan for på denne måten å vise utformingen av veggelementenes overflater:

Du må ta alle elementer fra kolonnen "Layer" med verdien "wall". Ta disse ID-ene som en liste fra "ID"-kolonnen og finn disse ID-ene fra hele datarammen i "ParentID"-kolonnen. Alle elementene er linjer som er kombinert til én enkelt polylinje. Hver linje har en annen x, y-geometri for det første punktet i "Point"-kolonnen. Du må ta hver polylinje etter tur og fra punktet 0,0 horisontalt tegne lengden på hvert segment fra polylinjen. lengden på hvert segment av polylinjen til en linje. Tegn deretter nøyaktig de samme linjene bare 3000 høyere, og koble alle punktene til ett plan.

- LLM-responsen vil sende ut kode som lar deg plotte veggtegninger i planet:

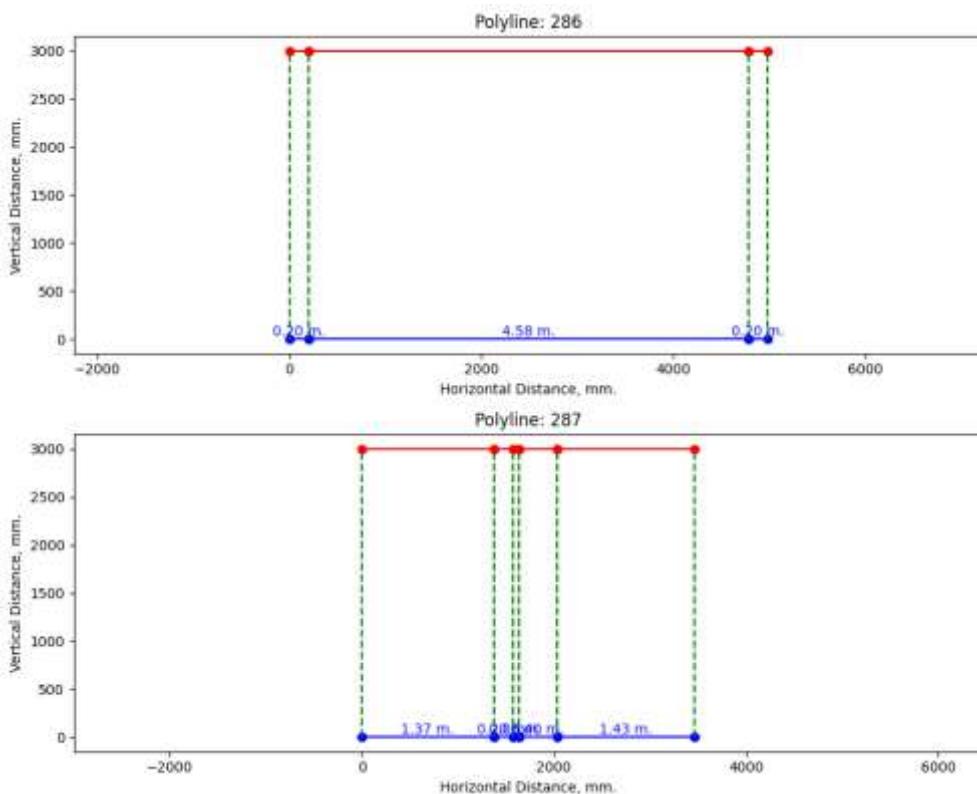


Fig. 6.4-10 Vi gjør om hver polylinje ved hjelp av ledetekster til en layout som visualiserer veggplanene direkte i LLM-chatten.

- La oss nå gå fra 2D-projeksjon til 3D - modellering av veger fra flate linjer ved å koble sammen de øverste og nederste lagene av polylinjer:

Visualiser veggelementer i 3D, ved å koble sammen polylinjer i høydene $z = 0$ og $z = 3000$ mm. For å lage en lukket geometri som representerer veggene i bygningen. Bruk Matplotlib 3D-grafverktøy.

- LLM vil generere en interaktiv 3D -graf der hver polylinje vil bli representert som et sett med plan. Brukeren vil kunne bevege seg fritt mellom elementene med en datamus og utforske modellen i 3D-modus ved å kopiere koden fra chatten til IDE-en:

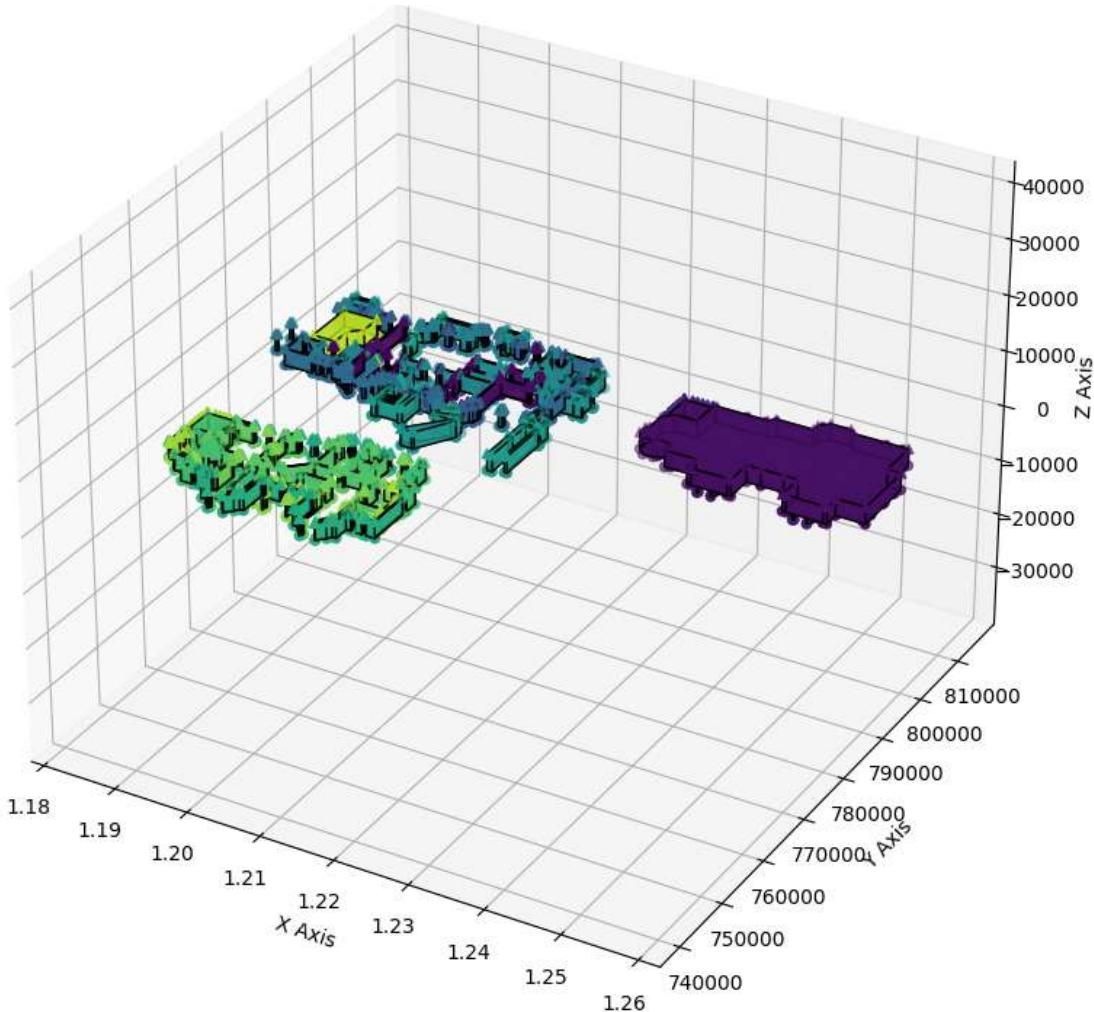


Fig. 6.4-11 LLM bidro til å lage kode [129] for å visualisere flate tegningslinjer i en 3D-visning på som kan utforskes i 3D-visningsprogrammet i IDE.

For å bygge opp en logisk og reproducibel Pipeline - fra den første konverteringen og innlastingen av DWG -filen til det endelige resultatet - anbefales det å kopiere den genererte LLM -kodeblokken til IDE etter hvert trinn. På denne måten kan du ikke bare sjekke resultatet i chatten, men også kjøre det i utviklingsmiljøet ditt umiddelbart. På denne måten kan du bygge opp prosessen sekvensielt, feilsøke og tilpasse den etter behov.

Du finner den komplette Pipelinefor alle fragmentene (figurene 6.4-8 til figurene 6.4-11) sammen med eksempler på spørninger på Kaggle-plattformen.com ved å søke på "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. På Kaggle kan du ikke bare se koden og spørringene som brukes, men også kopiere og teste hele rørledningen med de originale DWG-datarammene i skyen gratis uten å måtte

installere ekstra programvare eller selve IDE-en.

Metoden som presenteres i dette kapittelet, gjør det mulig å helautomatisere kontroll, behandling og generering av dokumenter basert på DWG -prosjekter. Den utviklede pipelinen eigner seg både for behandling av individuelle tegninger og for batchbehandling av titalls, hundrevis og tusenvis av DWG-filer med automatisk generering av nødvendige rapporter og visualiseringer for hvert prosjekt.

Prosesssen kan organiseres på en sekvensiell og transparent måte: Først konverteres dataene fra CAD -filen automatisk til XLSX-format, deretter lastes de inn i en dataramme, etterfulgt av gruppering, kontroll og resultatgenerering - alt dette implementeres i en enkelt Jupyter-notatbok eller Python -skript, i et hvilket som helst populært IDE. Om nødvendig kan prosessen enkelt utvides ved hjelp av integrasjon med systemer for styring av prosjektdokumentasjon: DAK-filer kan hentes automatisk i henhold til spesifiserte kriterier, resultatene kan returneres tilbake til lagringssystemet, og brukerne kan varsles når resultatene er klare - via e-post eller messengers.

Ved å bruke LLM chats og agenter til å arbeide med designdata reduserer du avhengigheten av spesialiserte CAD -programmer og kan utføre analyser og visualisering av arkitektoniske design uten behov for manuell interaksjon med grensesnittet - uten museklikk og uten å måtte huske kompleks menynavigering.

For hver dag som går, vil byggebransjen høre mer og mer om LLM, granulære strukturerte data, DataFrames og kolonnedbaser. Enhetlige todimensjonale DataFrames dannet av ulike databaser og CAD-formater, vil være det ideelle drivstoffet for moderne analyseverktøy som håndteres aktivt av spesialister i andre bransjer.

Selv automatiseringsprosessen vil bli betydelig forenklet - i stedet for å studere API for lukkede nisjeprodukter og skrive komplekse skript for å analysere eller transformere parametere, vil det nå være nok å formulere en oppgave i form av et sett med individuelle tekstkommandoer, som vil bli brettet inn i den nødvendige Pipeline eller Workflow-prosessen for det nødvendige programmeringsspråket, som kjører gratis på nesten alle enheter. Man trenger ikke lenger å vente på nye produkter, formater, plug-ins eller oppdateringer fra CAD- (BIM-) verktøyleverandører. Ingeniører og byggmenn vil kunne arbeide selvstendig med data ved hjelp av enkle, gratis og lettforståelige verktøy, assistert av LLM chatter og agenter.

Neste skritt: fra lukkede formater til åpne data

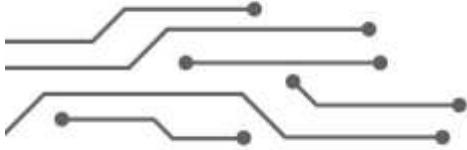
Når man skal jobbe med fremtidens designdata, er det lite sannsynlig at noen virkelig trenger å forstå de geometriske kjernene i proprietære verktøy eller lære seg hundrevis av inkompatible formater som inneholder den samme informasjonen. Men uten å forstå hvorfor det er viktig å gå over til åpne strukturerte data, er det vanskelig å argumentere for bruk av nye gratisverktøy, åpne data og tilnærminger som neppe vil bli fremmet av programvareleverandører.

I dette kapittelet har vi diskutert de viktigste egenskapene ved CAD (BIM), begrensningene og mulighetene som ligger i dem, og at ingeniører og designere, til tross for leverandørenes løfter i markedsføringen, hver eneste dag står overfor problemer med å hente ut, overføre og analysere designinformasjon. Å forstå arkitekturen i disse systemene og lære om alternative tilnærmingar - basert på åpne formater og automatisering gjennom LLM - kan gjøre livet mye enklere selv for en enkelt fagperson, for ikke å snakke

om bedrifter. For å oppsummere denne delen er det verdt å fremheve de viktigste praktiske stegene som vil hjelpe deg med å bruke tilnærmingene som er omtalt i denne delen, i dine daglige oppgaver:

- Utvid verktøykassen din for å jobbe med prosjektdata
 - Utforsk tilgjengelige plugin-moduler og verktøy for å hente ut data fra CAD - (BIM-) -systemene du bruker
 - Utforsk tilgjengelige SDK-er og API-er som gjør det mulig å automatisere datauttrekk fra lukkede formater uten å måtte åpne spesialisert programvare manuelt
 - Beherske grunnleggende ferdigheter i å arbeide med åpne, ikke-parametriske geometriformater (OBJ, glTF, USD, DAE) og tilhørende open source-biblioteker
 - Prøv å finne et system for lagring av prosjektmetadata som er adskilt fra geometrien utenfor CAD (BIM) for å forenkle analyse og integrering med andre systemer
 - Bruk LLM til å automatisere datakonvertering mellom formater
- Lag dine egne prosesser for håndtering av prosjektinformasjon
 - Begynne å beskrive oppgaver og modelleringskrav ved hjelp av parametere og deres verdier i enkle og strukturerte formater
 - Opprett et personlig bibliotek med skript eller kodeblokker for ofte utførte operasjoner
- Fremme bruken av åpne standarder i arbeidet ditt
 - Inviter kolleger og partnere til å dele data i åpne formater som ikke er begrenset av programvareleverandørens økosystem
 - Demonstrasjone fordelene ved å bruke strukturerte data ved hjelp av konkrete eksempler
 - Ta initiativ til diskusjoner om problemer med lukkede formater og mulige løsninger

Selv om du ikke kan endre bedriftens policy når det gjelder CAD - (BIM-) -plattformer, vil en personlig forståelse av prinsippene for arbeid med prosjektdata i åpne formater gi deg mulighet til å effektivisere arbeidet ditt betaktelig. Ved å lage dine egne verktøy og metoder for å hente ut og transformere data fra ulike formater kan du ikke bare optimalisere arbeidsflyten, men også få fleksibilitet til å omgå begrensningene i standard programvareløsninger.



VII DEL

DATADREVET BESLUTNINGSTAKING, ANALYSE, AUTOMATISERING OG MASKINLÆRING

Den syvende delen fokuserer på dataanalyse og prosessautomatisering i byggebransjen. Her diskuteres hvordan data blir grunnlaget for beslutningstaking, og prinsippene for visualisering av informasjon for effektiv analyse forklares. Nøkkelindikatorer (KPI), metoder for å evaluere avkastningen på investeringen (ROI) og oppretting av dashboard for prosjektovervåking beskrives i detalj. ETL-prosesser (Extract, Transform, Load) og automatisering av disse ved hjelp av pipelines (Pipeline) for å omdanne ulike data til strukturert informasjon som kan analyseres, blir viet spesiell oppmerksomhet. Arbeidsflytorkestreringsverktøy som Apache Airflow, Apache NiFi og n8n, som gjør det mulig å bygge automatiserte datarørledninger uten inngående programmeringskunnskaper, blir omtalt. Store språkmodeller (Large Language Models, LLM) og bruken av dem for å forenkle dataanalyse automatisere rutineoppgaver spiller en viktig rolle.

KAPITTEL 7.1.

DATAANALYSE OG DATADREVET BESLUTNINGSTAKING

Etter å ha samlet inn, strukturert, renset og verifisert informasjonen, har det oppstått et sammenhengende og analyserbart datasett. De foregående delene av boken har tatt for seg systematisering og strukturering av heterogene kilder - fra PDF-dokumenter og møtereferater til CAD-modeller og geometriske data. Prosessen med å kontrollere og tilpasse informasjonen til kravene fra ulike systemer og klassifikatorer, og eliminere duplikater og inkonsekvenser, er beskrevet i detalj.

Alle beregningene som utføres på disse dataene (tredje og fjerde del av boken) - fra enkle transformasjoner til beregninger av tids-, kostnads- og ESG-indikatorer (femte del) - er aggregerte analyseoppgaver. De danner grunnlaget for å forstå et prosjekts nåværende tilstand, vurdere parameterne og deretter ta beslutninger. Som et resultat av beregningene forvandles dataene fra et sett med ulike poster til en håndterbar ressurs som kan gi svar på viktige forretningsspørsmål.

I de foregående kapitlene har vi beskrevet datainnsamlings- og kvalitetskontrollprosessene for bruk i typiske forretningscase og prosesser som er spesifikke for byggebransjen. Analyse i denne sammenhengen ligner på mange måter på anvendelser i andre bransjer, men har en rekke særtrekk.

I de følgende kapitlene beskrives dataanalyseprosessen i detalj, inkludert automatiseringstrinnene fra den første informasjonsinnehenting og transformasjonen til den påfølgende overføringen til målsystemer og -dokumenter. Først presenteres en teoretisk del med fokus på utvalgte aspekter ved datadrevet beslutningstaking. Deretter, i de påfølgende kapitlene, presenteres den praktiske delen knyttet til automatisering og konstruksjon av ETL -Pipeline.

Data som en ressurs i beslutningsprosessen

Datadrevet beslutningstaking er ofte en iterativ prosess som begynner med systematisk innsamling av informasjon fra en rekke ulike informasjonskilder. Som i et naturlig kretsløp faller enkelte dataelementer og hele informasjonssystemer gradvis ned i jorden - og samler seg i bedriftens informasjonslagre (figur 1.3-2). Over tid blir disse dataene, i likhet med nedfallne blader og greiner, omdannet til verdifullt materiale. Myelet av dataingeniører og analyticikere organiserer og forbereder informasjon for fremtidig bruk og forandler nedfallne data og systemer til verdifull kompost, slik at nye skudd og nye systemer kan vokse frem (fig. 1.2-5).

Trender i utbredt bruk av analyseverktøy i ulike bransjer, markerer begynnelsen på en ny æra der arbeid med data blir grunnlaget for profesjonell aktivitet (fig. 7.1-1). Det er viktig at fagfolk i byggebransjen tilpasser seg disse endringene og er klare til å bevege seg inn i en ny æra - data- og analysetiden

Manuell flytting av data mellom tabeller og manuelle beregninger er i ferd med å bli en saga blott, til fordel for automatisering, dataflytanalyse, analyse og maskinlæring. Disse verktøyene er i ferd med å bli sentrale elementer i moderne beslutningsstøttesystemer.

I McKinseys bok "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] siteres en undersøkelse gjennomført i 2022 med 1330 toppledere fra ulike regioner, bransjer og funksjonsområder. I rapporten "Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] refereres det til en undersøkelse som ble gjennomført i 2022 blant 1330 toppledere fra ulike regioner, bransjer og funksjonsområder. Ifølge resultatene bruker 70% av lederne avanserte analyser til å generere sine egne ideer, og 50% implementerer kunstig intelligens for å forbedre og automatisere beslutningsprosesser.



Fig. 7.1-1 Dataanalyse og analyse er det viktigste verktøyet for å øke beslutningshastigheten i en bedrift.

Dataanalyse, som mycelium, trenger inn i humusen av tidligere beslutninger, bidrar til å knytte sammen individuelle systemer og veileder ledere til verdifull innsikt. Denne kunnskapen, som næring fra trærne i datasystemene, gir næring til nye beslutninger i bedriften, noe som fører til effektiv endring og vekst av kvalitetsinformasjon, som nye skudd og spirer som springer ut av rik og sunn jord (figur 1.2-5).

Tallene har en viktig historie å fortelle. De stoler på at du gir dem en klar og overbevisende stemme [131].

- Stephen Few, ekspert på datavisualisering

I mellomstore og små bedrifter er arbeidet med å hente ut og klargjøre informasjon for videre analyse i dag en ekstremt arbeidsintensiv prosess (figur 7.1-2), som kan sammenlignes med 1700-tallets kullgruvedrift. Inntil nylig var arbeidet med datautvinning og -forberedelse forbeholdt eventyrere som jobbet i en høyt spesialisert nisje med et lite og begrenset sett med verktøy for å jobbe med ulike typer data fra ustukturerte, løst strukturerede, blandede og lukkede kilder.

Beslutningstakere og ledere er ofte uerfarne med heterogene data og systemer, men må likevel ta beslutninger basert på dem. Som et resultat av dette har datadrevet beslutningstaking i den moderne byggebransjen de siste tiårene i mindre grad lignet en automatisert prosess og mer på det manuelle flerdagersarbeidet til en gruvearbeider i de tidlige kullgruvene.

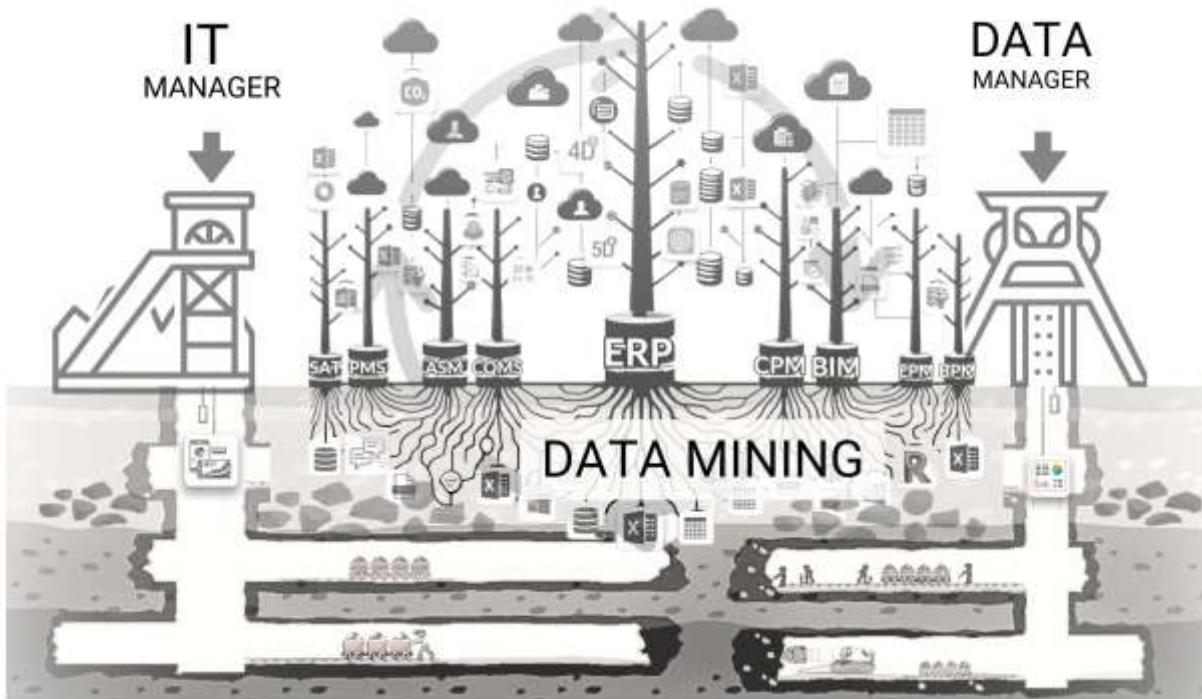


Fig. 7.1-2 I datautvinningsprosessen går ekspertene gjennom en kompleks prosess med klargjøring av data - fra rensing til strukturering for senere analyse.

Selv om moderne metoder for datautvinning i byggebransjen utvilsomt er mer avanserte enn de primitive teknikkene som ble brukt av gruvearbeidere på 1100-tallet, er det fortsatt en kompleks og risikofylt oppgave som krever betydelige ressurser og ekspertise som bare store selskaper har råd til. Prosessene med å hente ut og analysere data fra den akkumulerte arven fra tidligere prosjekter har inntil nylig hovedsakelig blitt utført av store teknologisk avanserte selskaper som har samlet inn og lagret data konsekvent i flere tiår.

Tidligere var det teknologisk modne selskaper som hadde samlet inn data i flere tiår, som hadde den ledende rollen innen analyse. I dag er situasjonen i ferd med å endre seg: Tilgangen til data og databehandlingsverktøy er i ferd med å demokratiseres - tidligere komplekse løsninger er nå gratis tilgjengelig for alle.

Ved å bruke analyser kan bedrifter ta mer nøyaktige og informerte beslutninger i sanntid. Følgende casestudie illustrerer hvordan historiske data kan bidra til å ta økonomisk fornuftige beslutninger:

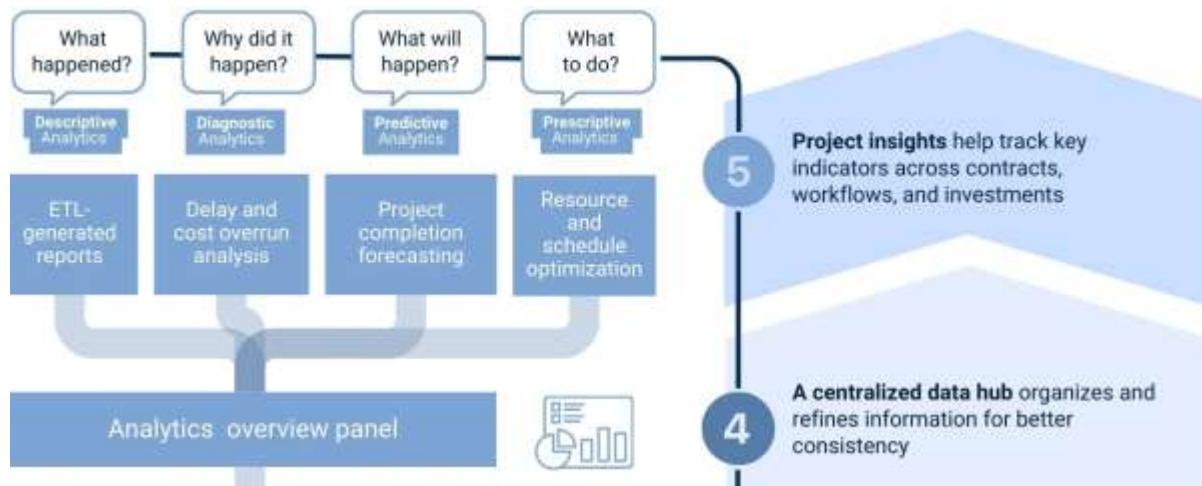
- ☛ **Prosjektleder** - "Nå er gjennomsnittsprisen på betong i byen 82€ /m³, vi har 95 €/m³ i vårt estimat."
- ☛ **Kalkulatør** - "På tidligere prosjekter var overskridelsen på rundt 15%, så jeg trakk meg tilbake."
- ☛ **Dataansvarlig eller kontrollingeniør på kundesiden** - "La oss se på analysene for de tre siste anbudene."

Etter å ha analysert DataFrame fra tidligere prosjekter, får vi:

- **Gjennomsnittlig faktisk innkjøpspris:** 84,80 /m³
- **Gjennomsnittlig overskridelsesprosent:** +4,7

💡 Anbefalt hastighet i estimatet: ~ 85 /m³€

En slik beslutning vil ikke lenger være basert på subjektive følelser, men på spesifikk historisk statistikk, noe som bidrar til å redusere risikoen og øke anbudets validitet. Analyse av data fra tidligere prosjekter blir en slags "organisk gjødsel" som gir grobunn for nye, mer nøyaktige løsninger.



Figur 7.1-3 Dataanalyse gir svar på tre viktige spørsmål: hva som skjedde, hvorfor det skjedde, og hva som bør gjøres videre.

Beslutningstakere og ledere står ofte overfor behovet for å arbeide med heterogene data og systemer uten tilstrekkelig teknisk ekspertise. I slike situasjoner er visualisering, som er et av de første og viktigste trinnene i analyseprosessen, et viktig hjelpemiddel for å forstå data. Det gjør det mulig å presentere informasjon i en visuell og forståelig form.

Visualisering av data: nøkkelen til forståelse og beslutningstaking

I dagens byggebransje, der prosjektdata kjennetegnes av kompleksitet og struktur på flere nivåer, spiller visualisering en nøkkelrolle. Visualisering av data gjør det mulig for prosjektledere og ingeniører å visualisere komplekse mønstre og trender som skjuler seg i store, heterogene datamengder.

Visualisering av data gjør det enklere å forstå statusen til et prosjekt: ressursallokering, kostnadstrender eller materialbruk. Grafer og diagrammer gjør kompleks og tørr informasjon tilgjengelig og forståelig, slik at du raskt kan identifisere viktige områder som trenger oppmerksomhet og oppdage potensielle problemer.

Visualisering av data gjør det ikke bare lettere å tolke informasjonen, det er også et avgjørende skritt i analyseprosessen og i en informert beslutningsprosess som bidrar til å besvare spørsmålene "hva skjedde?" og "hvordan skjedde det?" (figur 2.2-5). (Figur 2.2-5).

Grafikk er visuelle verktøy for å løse logiske problemer [132].

- Jacques Bertin, "Grafikk og grafisk informasjonsbehandling"

Før de tar viktige beslutninger, er det mer sannsynlig at prosjektledere bruker visuelle fremstillinger av data i stedet for tørre og vanskelig tolkbare tall fra regneark eller tekstmeldinger.

Data uten visualisering er som byggematerialer som ligger strødd på måfå på en byggeplass: Potensialet deres er uklart. Det er først når de visualiseres tydelig, som et hus av murstein og betong, at verdien av dem blir klar. Før huset er bygget, er det umulig å si om materialhaugen vil bli en liten hytte, en luksusvilla eller en skyskraper.

Bedrifter har data fra ulike systemer (figur 1.2-4 til figur 2.1-10), økonomiske transaksjoner og omfattende tekstdata. Det er imidlertid ofte utfordrende å utnytte disse dataene til fordel for virksomheten. I slike situasjoner blir visualisering et viktig verktøy for å kommunisere betydningen av dataene, og det bidrar til å presentere informasjonen i formater som alle eksperter kan forstå, for eksempel i form av dashbord, grafer og diagrammer.

PwCs studie "What Students Need to Succeed in a Fast Changing Business World" (2015) fremhever [9] at vellykkede selskaper går lengre enn dataanalyse og aktivt bruker interaktive visualiseringsverktøy som grafer, infografikk og analytiske dashbord for å støtte beslutningstaking. Ifølge rapporten hjelper datavisualisering kundene med å forstå historien som dataene forteller gjennom grafer, diagrammer, instrumentpaneler og interaktive datamodeller.

Prosessen med å konvertere informasjon til visuelle grafiske former som diagrammer, grafer og diagrammer gjør det lettere for den menneskelige hjerne å forstå og tolke data (figur 7.1-4). På denne måten kan prosjektledere og analyticikere raskere vurdere komplekse scenarier og ta velbegrunnede beslutninger basert på visuelt gjenkjennelige trender og mønstre i stedet for intuisjon.

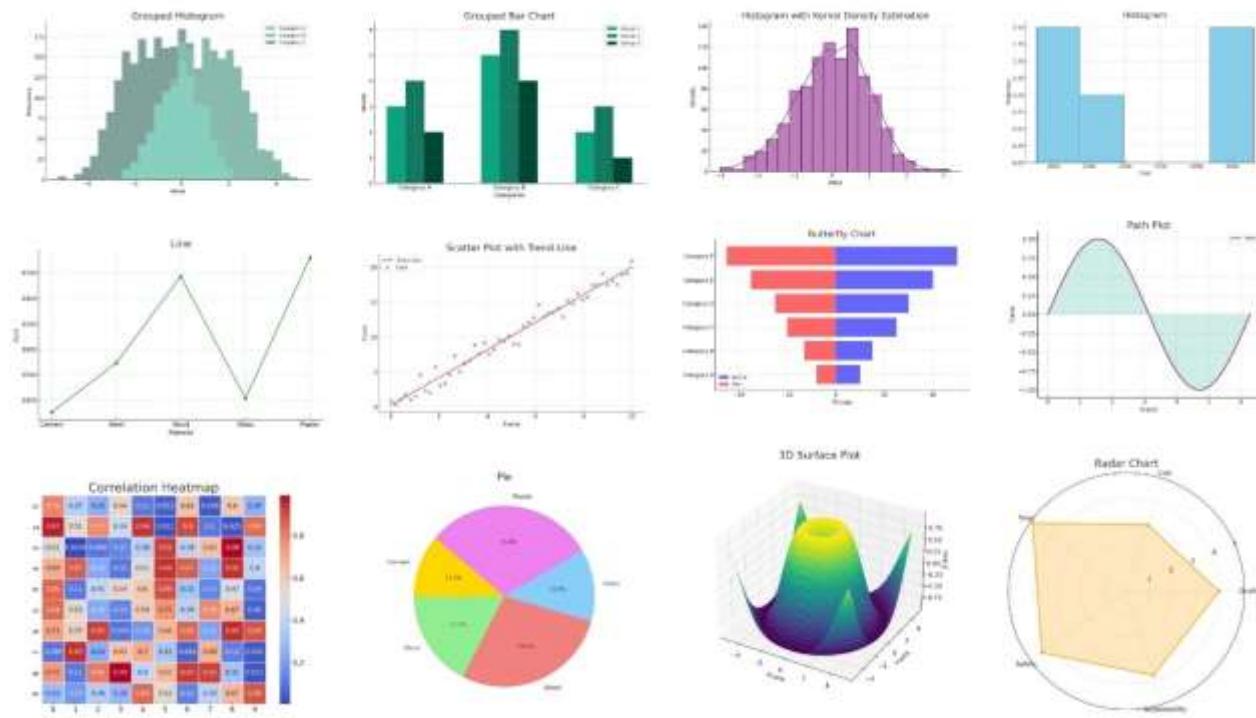


Fig. 7.1-4 Ulike typer visualisering er utviklet for å hjelpe menneskehjernen med å forstå og forstå den tørre tallinformasjonen bedre.

Hvordan man lager visualiseringer fra data, og hvordan man bruker ulike gratis visualiseringsbiblioteker, vil bli diskutert mer inngående i neste kapittel om ETL -prosesser.

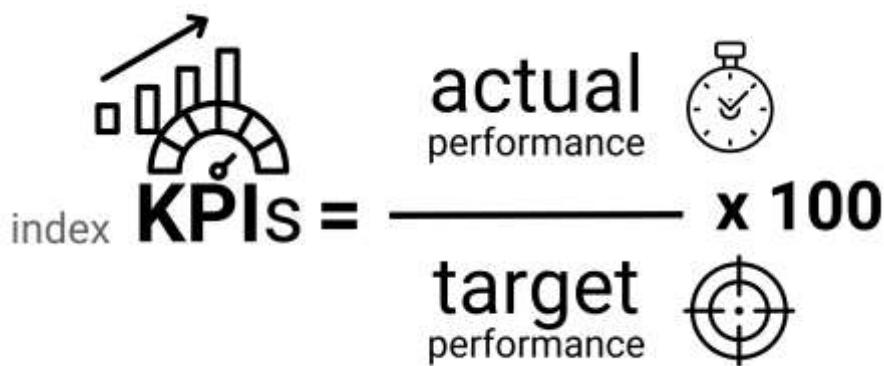
Visualisering er i ferd med å bli et integrert element i arbeidet med data i byggebransjen - det hjelper ikke bare med å "se" data, men også med å forstå betydningen av dem i forbindelse med ledelsesoppgaver. Men for at visualisering skal være virkelig nyttig, må man på forhånd avgjøre hva som skal visualiseres, og hvilke måleparametere som virkelig er viktige for å vurdere prosjektets ytelse. Det er her prestasjonsmålinger som KPI-er og ROI kommer inn i bildet. Uten dem risikerer selv de flotteste dashbord å bare bli "informasjonsstøy".

KPI-er og ROI

I dagens byggebransje spiller styring av resultatindikatorer (KPI og ROI) og visualisering av disse gjennom rapporter og dashbord en nøkk爾olle når det gjelder å forbedre produktiviteten og effektiviteten i prosjektstyringen.

Som i alle andre virksomheter er det også i byggebransjen viktig å definere klare måleparametere for suksess, avkastning på investeringene og prestasjoner. Når en datadrevet organisasjon skal innhente data om ulike prosesser, må den først lære seg å identifisere viktige **KPI-er (Key Performance Indicators)** - kvantitative mål som gjenspeiler i hvilken grad strategiske og operasjonelle mål blir nådd.

For å beregne en KPI bruker man vanligvis en formel (figur 7.1-5) som inkluderer faktiske og planlagte indikatorer. For å beregne en individuell KPI for et prosjekt, en medarbeider eller en prosess, dividerer man for eksempel den faktiske ytelsen med den planlagte ytelsen og multipliserer resultatet med 100%.



Figur 7.1-5 KPIer brukes til å måle hvor vellykket et prosjekt eller en prosess er når det gjelder å nå viktige mål.

På avdelingsnivå kan mer detaljerte KPI-er brukes metrics:

- **Tidspunkt for viktige milepæler** (fundamentering, installasjon, ferdigstillelse) - gjør det mulig å kontrollere at arbeidsplanene overholdes.
- **Prosentandel av materialoverskridelser** - bidrar til å styre innkjøp og minimere svinn.
- **Antall uplanlagte maskinstopp** - påvirker produktivitet og kostnader.

Hvis man velger feil måleparametere, kan det føre til feilaktige "hva skal jeg gjøre?"-beslutninger (figur 2.2-5). Hvis et selskap for eksempel bare fokuserer på kostnad per kvadratmeter, men ikke tar hensyn til kostnadene ved ombygginger, kan besparelser på materialer føre til dårligere kvalitet og høyere kostnader i fremtidige prosjekter.

Når man setter opp mål, er det viktig å være tydelig på hva som skal måles. Vage formuleringer fører til feilaktige konklusjoner og vanskelig gjør kontroll. La oss se på eksempler på vellykkede og mislykkede KPI-er i bygg- og anleggsbransjen.

Gode KPI-er:

- ❑ "Innen utgangen av året skal vi redusere andelen oppussingsarbeid med 10 prosent."
- ❑ "Øke hastigheten på fasademonitasje med 15 prosent uten at det går på bekostning av kvaliteten innen neste kvartal"
- ❑ "Redusere nedetiden på maskineriet med 20% ved å optimalisere arbeidsplanene innen utgangen av året"

Disse målene er tydelig målbare, med spesifikke verdier og tidsrammer.

Dårlige KPI-er:

- ❑ "Vi skal bygge raskere" (Hvor mye raskere? Hva betyr "raskere"?).
- ❑ "Vi skal forbedre kvaliteten på betongarbeidet" (Hvordan måles egentlig kvalitet?)

-  "Vi skal forbedre samspillet med entreprenøren på byggeplassen" (Hvilke kriterier vil vise forbedring?)

En god KPI er en som kan måles og vurderes objektivt. I byggebransjen er dette spesielt viktig, for uten klare indikatorer er det umulig å overvåke prestasjoner og oppnå stabile resultater.

I tillegg til KPI finnes det en ytterligere måleenhet for å vurdere effektiviteten av investeringer: **ROI (Return on Investment)** - en avkastningsindikator som gjenspeiler forholdet mellom fortjeneste og investerte midler. ROI bidrar til å vurdere om innføringen av nye metoder, teknologier eller verktøy er berettiget: fra digitale løsninger og automatisering (se f.eks. figur 7.3-2 til bruk av nye byggematerialer. Denne indikatoren bidrar til å ta informerte beslutninger om videre investeringer basert på deres reelle innvirkning på virksomhetens lønnsomhet

I forbindelse med styring av byggeprosjekter kan ROI (return on investment) brukes som en av nøkkelindikatorene (KPI) hvis bedriftens mål er å måle avkastningen på investeringen i et prosjekt, en teknologi eller en prosessforbedring. Hvis for eksempel en ny byggeledelsesteknikk er under implementering, kan ROI vise hvor mye den har forbedret lønnsomheten.

Regelmessig måling av KPI-er og ROI-er basert på data som sammles inn fra ulike kilder, for eksempel materialforbruk, arbeidstimer og kostnader, gjør det mulig for prosjektledelsen å styre ressursene effektivt og ta raske beslutninger. Ved å lagre disse dataene på lang sikt kan fremtidige trender analyseres og prosesser optimaliseres.

Ulike diagrammer og grafer brukes til å visualisere KPI-er, ROI-er og andre måleparametere og kombineres vanligvis i dashbord.

Dashbord og instrumentpaneler: Visualisering av måltall for effektiv ledelse

En rekke diagrammer og grafer brukes til å visualisere indikatorer og måleparametere, som vanligvis kombineres i datautstillings vinduer og dashbord. Disse instrumentbordene gir en sentralisert oversikt over statusen til et prosjekt eller deler av et prosjekt, og viser nøkkelindikatorer (helst i sanntid). Oppdaterte og kontinuerlig oppdaterte dashbord gjør det mulig for teamet å reagere raskt på endringer.

Dashbord er verktøy som visualiserer kvantitative vurderinger og gjør dem lett tilgjengelige og forståelige for alle prosjektdeltakerne.

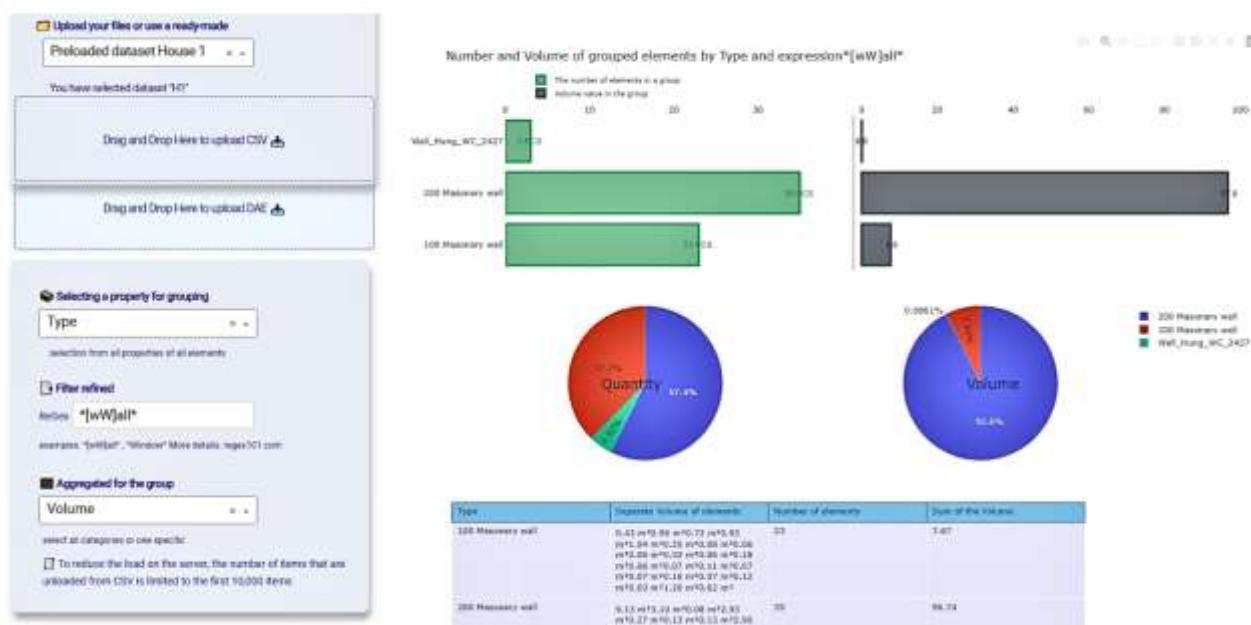


Fig. 7.1-6 Administrere KPI-er og visualisere dem ved hjelp av instrumentpaneler er nøkkelen til å forbedre prosjektets produktivitet og effektivitet.

Her er noen eksempler på populære verktøy du kan bruke til å lage dashbord:

- **Power BI** er et verktøy fra Microsoft for å lage interaktive rapporter og dashbord.
- **Tableau og Google Data Studio** er kraftige verktøy for å visualisere data og lage dashbord uten å måtte skrive kode.
- **Plotly** (figur 7.1-6, figur 7.2-12) er et bibliotek for å lage interaktive grafer, og Dash er et rammeverk for å lage webapplikasjoner for analyse av data. De kan brukes i kombinasjon for å lage interaktive dashbord.
- **Mange Python-biblioteker** (Figur 7.2-9 - Figur 7.2-11) - Python har mange åpen kildekode- og gratisbiblioteker for datavisualisering, for eksempel Matplotlib, Seaborn, Plotly, Bokeh og andre. Disse kan brukes til å lage grafer og integrere dem i en webapplikasjon ved hjelp av rammeverk som Flask eller Django.
- **JavaScript Libraries**: lar deg lage interaktive dashbord ved hjelp av Open Source JavaScript-biblioteker som D3.js eller Chart.js og integrere dem i en webapplikasjon.

For å kunne vurdere KPI-er og lage dashbord trenger du oppdaterte data og en klar tidsplan for innsamling og analyse av informasjon.

Generelt danner KPI, ROI og dashbord i byggebransjen grunnlaget for en analytisk tilnærming til prosjektstyring. De bidrar ikke bare til å overvåke og evaluere den nåværende statusen, men gir også verdifull innsikt for fremtidige planleggings- og optimaliseringsprosesser - prosesser som er direkte avhengige av å tolke data og stille de riktige spørsmålene i tide.

Dataanalyse og kunsten å stille spørsmål

Datatolkningen er den siste fasen av analysen, der informasjonen gir mening og begynner å "snakke". Det er her svarene på de viktigste spørsmålene formuleres: "Hva skal vi gjøre?" og "Hvordan skal vi gjøre?" (figur 2.2-5). I denne fasen kan man oppsummere resultatene, identifisere mønstre, etablere årsakssammenhenger og trekke konklusjoner basert på visualisering og statistisk analyse.

Kanskje er tiden ikke langt unna da man vil innse at det er like nødvendig å kunne regne, tenke i gjennomsnitt, maksima og minima, som det nå er å kunne lese og skrive, for å bli en effektiv borger i en av de store, nye, komplekse verdensstatene som er i ferd med å utvikle seg [133].

- Samuel S. Wilkes, sitert i en presidenttale til American Statistical Association i 1951

Ifølge rapporten "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) utgitt av den britiske regjeringen [83] kan implementeringen av analyser data og kunstig intelligens (AI) forbedre prosjektstyringsprosessene betydelig, øke nøyaktigheten i tids- og kostnadsprognosør, samt redusere risiko og usikkerhet. I artikkelen fremheves det at offentlige organisasjoner som bruker avanserte analyseverktøy, oppnår bedre resultater i infrastrukturprosjekter.

Moderne byggevirksomhet som opererer i den fjerde industrielle revolusjonens sterkt konkurranseutsatte miljø med lave marginer, kan sammenlignes med militære operasjoner. Her er bedriftens overlevelse og suksess avhengig av hvor raskt den får tak i ressurser og kvalitetsinformasjon - og dermed av at den kan ta informerte beslutninger i tide (fig. 7.1-7).

Hvis datavisualisering er "etterretningen" som gir oversikt, er dataanalyse "ammunisjonen" som trengs for å handle. Den gir svar på spørsmålene "hva skal jeg gjøre?" og "hvordan skal jeg gjøre det?", og danner grunnlaget for å oppnå et konkurransefortrinn i markedet.

Analyse gjør ulike data om til strukturer og meningsfull informasjon som beslutningsgrunnlag.

Analytikernes og ledernes oppgave er ikke bare å tolke informasjon, men også å ta informerte beslutninger, identifisere trender, finne sammenhenger mellom ulike typer data og kategorisere dem i henhold til prosjektets mål og spesifikasjoner. Ved hjelp av visualiseringsverktøy og statistiske analysemetoder gjør de data til en strategisk ressurs for bedriften.

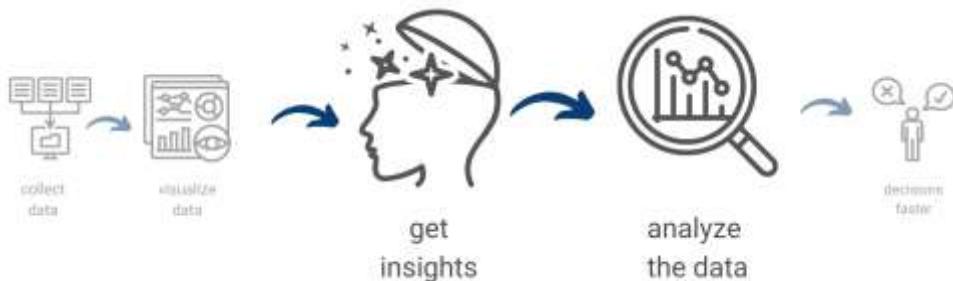


Fig. 7.1-7 Det er dataanalysentil syvende og sist gjør den innsamlede informasjonen til en kilde for beslutningstaking.

For å kunne ta velbegrunnede beslutninger i analyseprosessen, er det nødvendig å lære seg å formulere spørsmålene som stilles til dataene, på riktig måte. Kvaliteten på disse spørsmålene har direkte innvirkning på dybden i innsikten man oppnår, og dermed også på kvaliteten på ledelsesbeslutningene.

Fortiden eksisterer bare i den grad den er til stede i dagens nedtegnelser. Og hva disse nedtegnelsene representerer, bestemmes av spørsmålene vi stiller. Det finnes ingen annen historie enn denne [134].

- John Archibald Wheeler, fysiker 1982

Kunsten å stille dyptpløyende spørsmål og tenke kritisk er en viktig ferdighet når man jobber med data. De fleste har en tendens til å stille enkle, overfladiske spørsmål som det krever liten innsats å besvare. En ekte analyse begynner imidlertid med meningsfulle og gjennomtenkte spørsmål som kan avdekke skjulte sammenhenger og årsakssammenhenger i informasjon som kan være skjult bak flere lag av resonnementer.

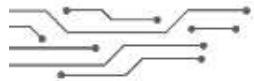
Ifølge studien "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] krever vellykket digital transformasjon investeringer i analysekapasitet, programmer for endringsledelse og tilpasning av forretningsmål til IT-initiativer. Bedrifter som skaper en datadrevet kultur, bør investere i dataanalysekapasitet og lansere programmer for endringsledelse for å innføre ny tenkning, atferd og arbeidsmåter.

Uten investeringer i utvikling av en analytisk kultur, forbedring av dataverktøy og opplæring av spesialister vil selskapene fortsatt risikere å ta beslutninger basert på utdatert eller ufullstendig informasjon - eller å stole på HiPPO-ledernes subjektive meninger (figur 2.1-9).

Erkjennelsen av relevansen og behovet for kontinuerlig oppdatering av analyser og dashbord fører uunngåelig til at ledelsen forstår viktigheten av å automatisere analytiske prosesser. Automatisering øker beslutningshastigheten, reduserer den menneskelige faktoren og sikrer datarelevansen. Med den eksponentielle veksten i informasjonsmengden blir hastighet ikke bare et konkurransefortrinn, men en nøkkelfaktor for bærekraftig suksess.

Automatisering av dataanalyse- og databehandlingsprosesser generelt er uløselig knyttet til temaet ETL

(Extract, Transform, Load). På samme måte som i automatiseringsprosessen må vi transformere data, blir data i ETL-prosessen hentet ut fra ulike kilder, transformert i henhold til de nødvendige kravene og lastet inn i målsystemene for videre bruk.



KAPITTEL 7.2.

DATAFLYT UTEN MANUELL INNSATS: HVORFOR ETL ER NØDVENDIG

ETL-automatisering: reduserte kostnader og raskere datahåndtering

Når nøkkelindikatorene (KPI-er) slutter å vokse til tross for at datamengden og teamstørrelsen øker, innser bedriftsledelsen uunngåelig behovet for å automatisere prosessene. Før eller senere blir denne erkjennelsen et incentiv til å sette i gang kompleks automatisering, der hovedmålet er å redusere kompleksiteten i prosessene, øke hastigheten på behandlingen og redusere avhengigheten av den menneskelige faktoren.

Ifølge McKinseys studie "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] får selskaper som bruker arkitekturen for strømming av data, en betydelig fordel fordi de kan analysere informasjon i sanntid. Strømmeteknologier muliggjør direkte analyse av sanntidsmeldinger og bruk av prediktivt vedlikehold i produksjonen ved å analysere sensordata i sanntid.

Prosessforenkling er automatisering, der tradisjonelle manuelle funksjoner erstattes av algoritmer og systemer.

Spørsmålet om automatisering, eller rettere sagt "minimering av menneskets rolle i databehandlingen", er en irreversibel og svært sensitiv prosess for alle selskaper. Spesialister på et hvilket som helst fagfelt kvier seg ofte for å avsløre sine arbeidsmetoder og finesser for andre optimalisatorer, fordi de er klar over risikoen for å miste jobben i et teknologisk miljø som er i rask utvikling.

Hvis du vil skaffe deg fiender, må du prøve å forandre ting [137].

- Woodrow Wilson, tale til en kongress for handelsmenn, Detroit, 1916

Til tross for de åpenbare fordelene med automatisering har mange bedrifter fortsatt en høy andel manuelt arbeid i den daglige driften, spesielt når det gjelder tekniske data. La oss se på et typisk eksempel på sekvensiell databehandling i slike prosesser for å illustrere dagens situasjon.

Manuell databehandling kan illustreres med et eksempel på interaksjon med informasjon hentet fra CAD-databaser. Tradisjonell databehandling ("manuell" ETL -prosess) i CAD (BIM)-avdelinger for å opprette attributtabeller eller lage dokumentasjon basert på prosjekteringsdata foregår i følgende rekkefølge (fig. 7.2-1):

1. **Manuell uttrekking (Extract):** Brukeren åpner prosjektet manuelt - ved å starte CAD-programmet (BIM) (fig. 7.2-1 trinn 1).
2. **Verifisering:** Neste trinn innebefatter vanligvis manuell kjøring av flere plugin-moduler eller hjelpeprogrammer for å klargjøre dataene og vurdere kvaliteten (fig. 7.2-1 trinn 2-3).
3. **Manuell transformasjon (Transform):** Etter klargjøring begynner databehandlingen, som krever manuell betjening av ulike programvareverktøy der dataene klargjøres for opplasting (fig. 7.2-1 trinn 4).

4. Manuell **opplasting (Load)**: manuell opplasting av konverterte data til eksterne systemer, dataformater og dokumenter (Fig. 7.2-1 trinn 5).

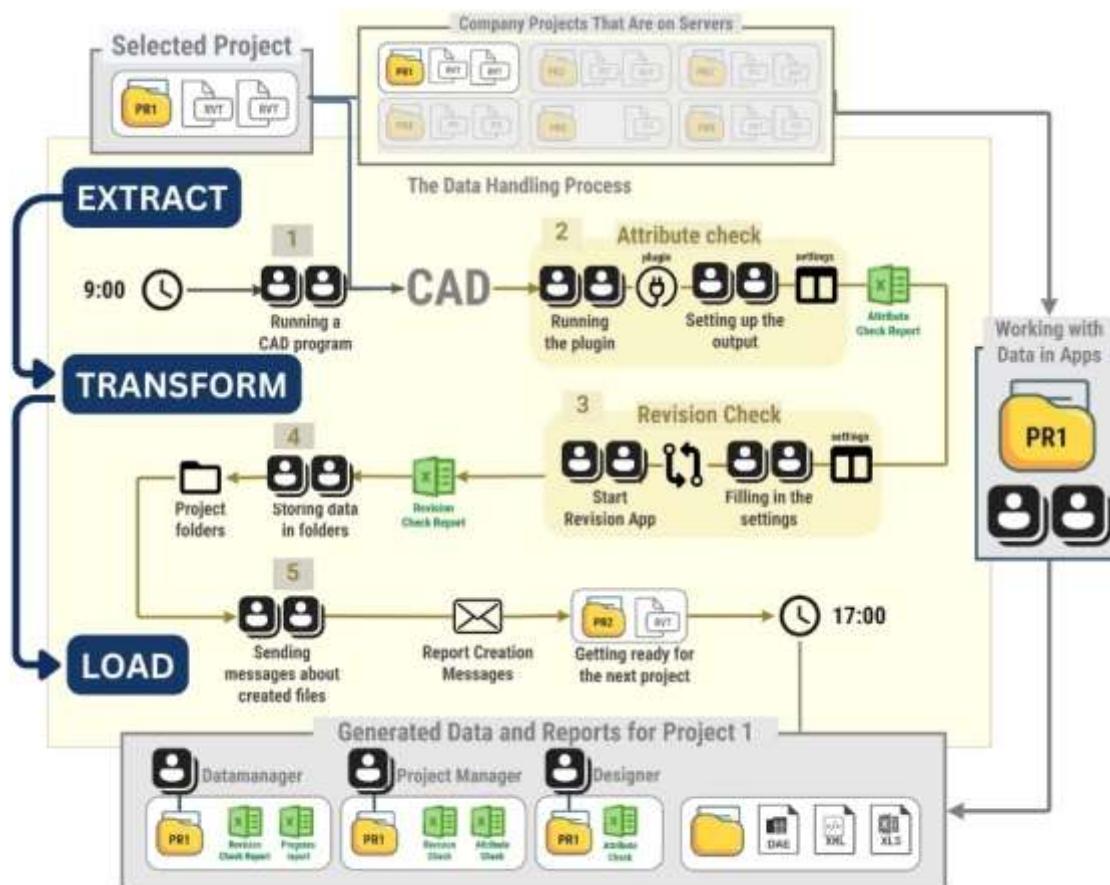


Fig. 7.2-1 Tradisjonell manuell ETL behandling er begrenset av den enkelte teknikers ønsker og fysiske evner.

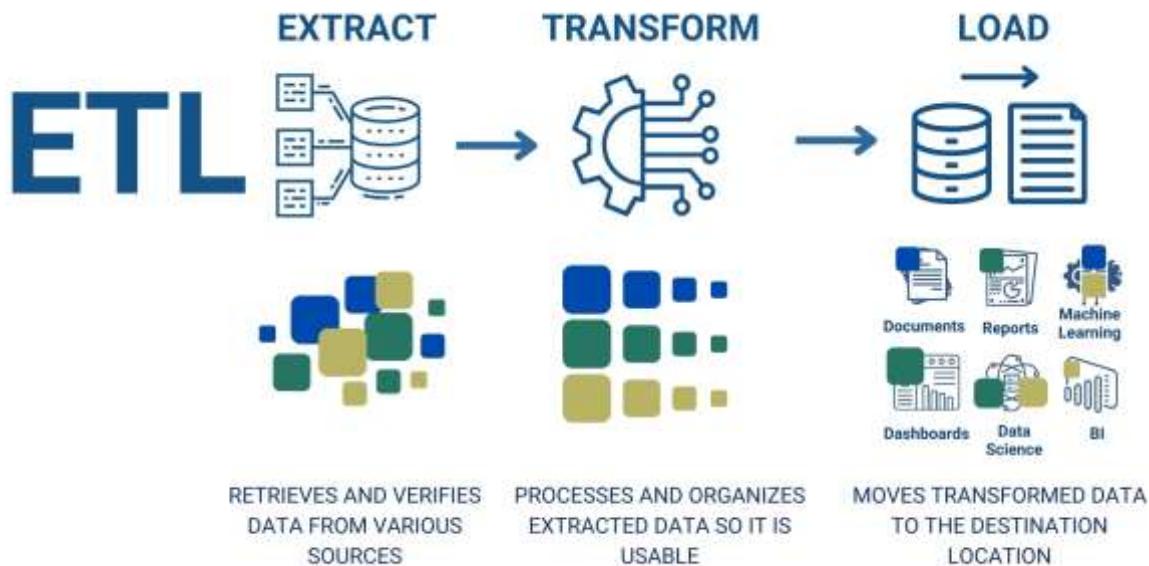
En slik arbeidsflyt er et eksempel på en klassisk ETL-prosess - ekstraksjon, transformasjon og lasting (ETL). I motsetning til andre bransjer, der automatiske ETL-rørledninger lenge har vært standard, er byggebransjen fortsatt dominert av manuelt arbeid, noe som forsinker prosessene og øker kostnadene

ETL (Extract, Transform, Load) er prosessen med å hente ut data fra ulike kilder, omforme dem til ønsket format og laste dem inn i målsystemet for videre analyse og bruk.

ETL er en prosess som omfatter tre hovedkomponenter i databehandlingen: Extract, Transform og Load (figur 7.2-2):

- **Trekk ut** - trekk ut data fra ulike kilder (filer, databaser, API).
- **Transform** - datarensing, aggregering, normalisering og logisk behandling.
- **Last inn** - last inn strukturert informasjon i et datavarehus, en rapport eller et BI-system.

Tidligere i boken har ETL-begrepet bare blitt berørt sporadisk: i forbindelse med konvertering av et ustrukturert, skannet dokument til et strukturert tabellformat (fig. 4.1-1), i forbindelse med formalisering av krav for å systematisere oppfatningen av både livs- og forretningsprosesser (fig. 4.4-20), og i forbindelse med automatisering av datavalidering og databehandling fra CAD-løsninger. La oss nå se nærmere på ETL i sammenheng med typiske arbeidsflyter.



Figur 7.2-2 ETL automatiserer repetitive databehandlingsoppgaver.

Manuell eller halvautomatisert ETL -prosess innebærer at en leder eller tekniker håndterer alle trinn manuelt - fra datainnsamling til rapportgenerering. En slik prosess tar mye tid, spesielt når arbeidstiden er begrenset (f.eks. 9:00 til 17:00).

Bedrifter søker ofte å løse problemet med lav effektivitet og lav hastighet ved å kjøpe modulbaserte integrerte løsninger (ERP, PMIS, CPM, CAFM, etc.), som deretter videreutvikles av eksterne leverandører og konsulenter. Men disse leverandørene og tredjepartsutviklerne blir ofte et kritisk avhengighetspunkt: Deres tekniske begrensninger påvirker direkte ytelsen til hele systemet og virksomheten som helhet, som beskrevet i de foregående kapitlene om proprietære systemer og formater. Problemene som oppstår på grunn av fragmentering og avhengighet, ble diskutert i detalj i kapittelet "Hvordan byggvirksomheter drukner i datakaos".

Hvis en bedrift ikke er klar til å implementere en stor modulær plattform fra en av leverandørene, begynner den å lte etter alternative måter å automatisere på. En av dem er å utvikle sine egne modulære, åpne ETL -transportører, der hvert trinn (ekstraksjon, transformasjon, validering, innlasting) implementeres som skript som kjøres etter en tidsplan.

I den automatiserte versjonen av den samme ETL-arbeidsflyten (fig. 7.2-1) ser arbeidsprosessen ut som en modulær kode som starter med å behandle data og oversette dem til en åpen strukturert form. Når de strukturerte dataene er mottatt, kjøres ulike skript eller moduler automatisk, på en planlagt basis, for å

kontrollere endringer, transformere og sende meldinger (fig. 7.2-3).

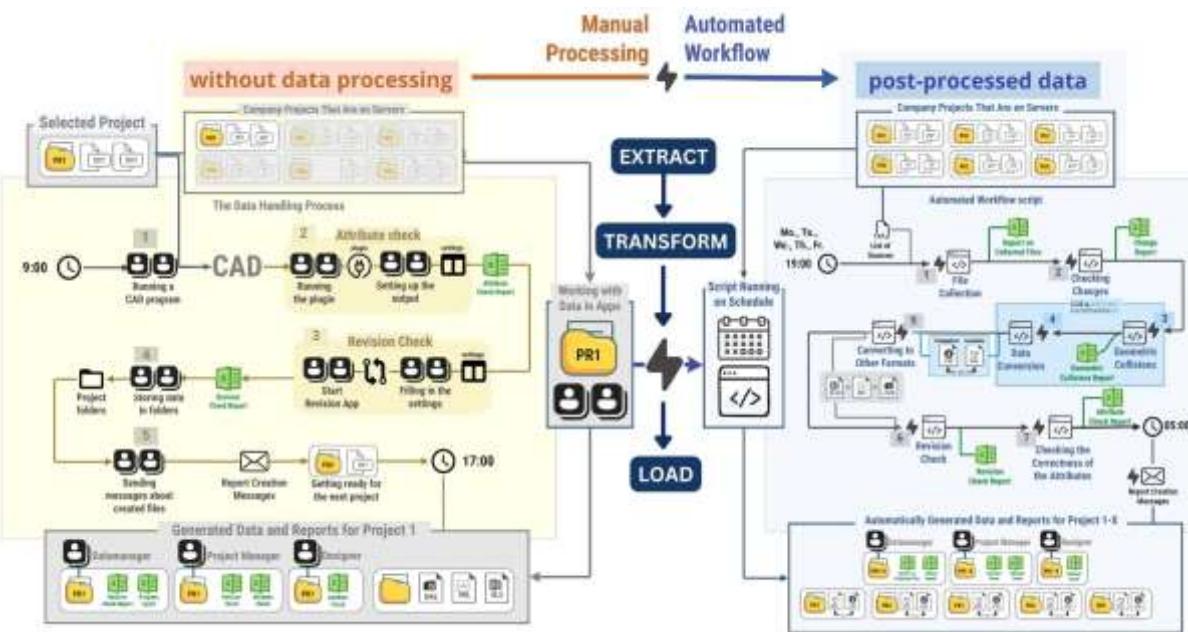


Fig. 7.2-3 Til venstre manuell maskinering, til høyre en automatisk prosess som, i motsetning til tradisjonell manuell maskinering, ikke er begrenset av brukerens evner.

I en automatisert arbeidsflyt forenkles databehandlingen ved hjelp av ET(L)-forbehandling av data: strukturering og forening.

I tradisjonelle behandlingsmetoder arbeider spesialister med data "som de er" - slik de hentes ut fra systemer eller programvare. I automatiserte prosesser, derimot, passerer dataene ofte først gjennom et ETL -payplane, der de bringes til en konsistent struktur og et format som er egnet for videre bruk og analyse.

La oss ta et praktisk ETL-eksempel, som demonstrerer valideringsprosessen for databatterier som er beskrevet i kapittelet "Validering av data og valideringsresultater" (figur 4.4-13). For å gjøre dette bruker vi Pandas-biblioteket sammen med LLM for automatiserte dataanalyse- og prosesseringss prosesser.

ETL Utdrag: datainnsamling

Den første fasen av ETL - Extract) - starter med å skrive kode for å samle inn datasett som skal kontrolleres og behandles videre. For å gjøre dette skanner vi alle mappene på produksjonsserveren, samler inn dokumenter med et bestemt format og innhold, og konverterer dem deretter til en strukturert form. Denne prosessen er beskrevet i detalj i kapitlene "Konvertering av ustukturerte og tekstlige data til strukturert form" og "Konvertering av CAD-data (BIM) til strukturert form" (fig. 4.1-1 - fig. 4.1-12).

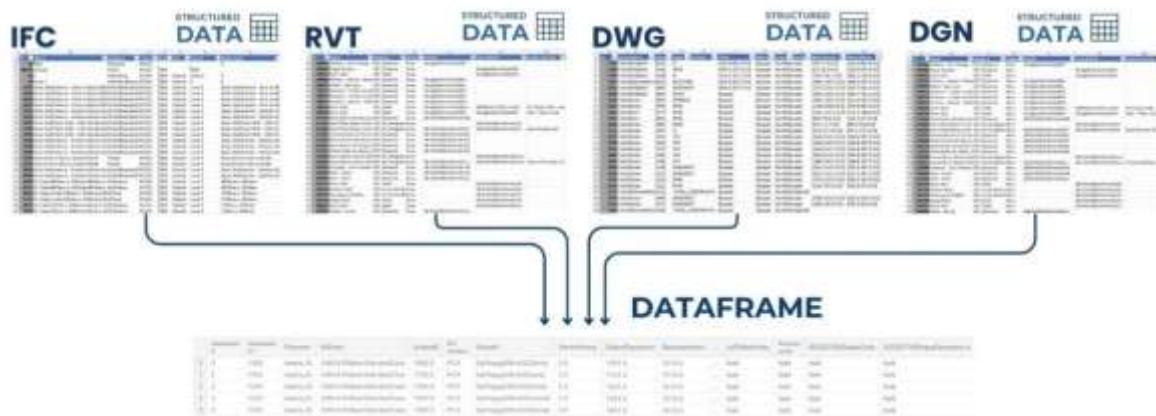
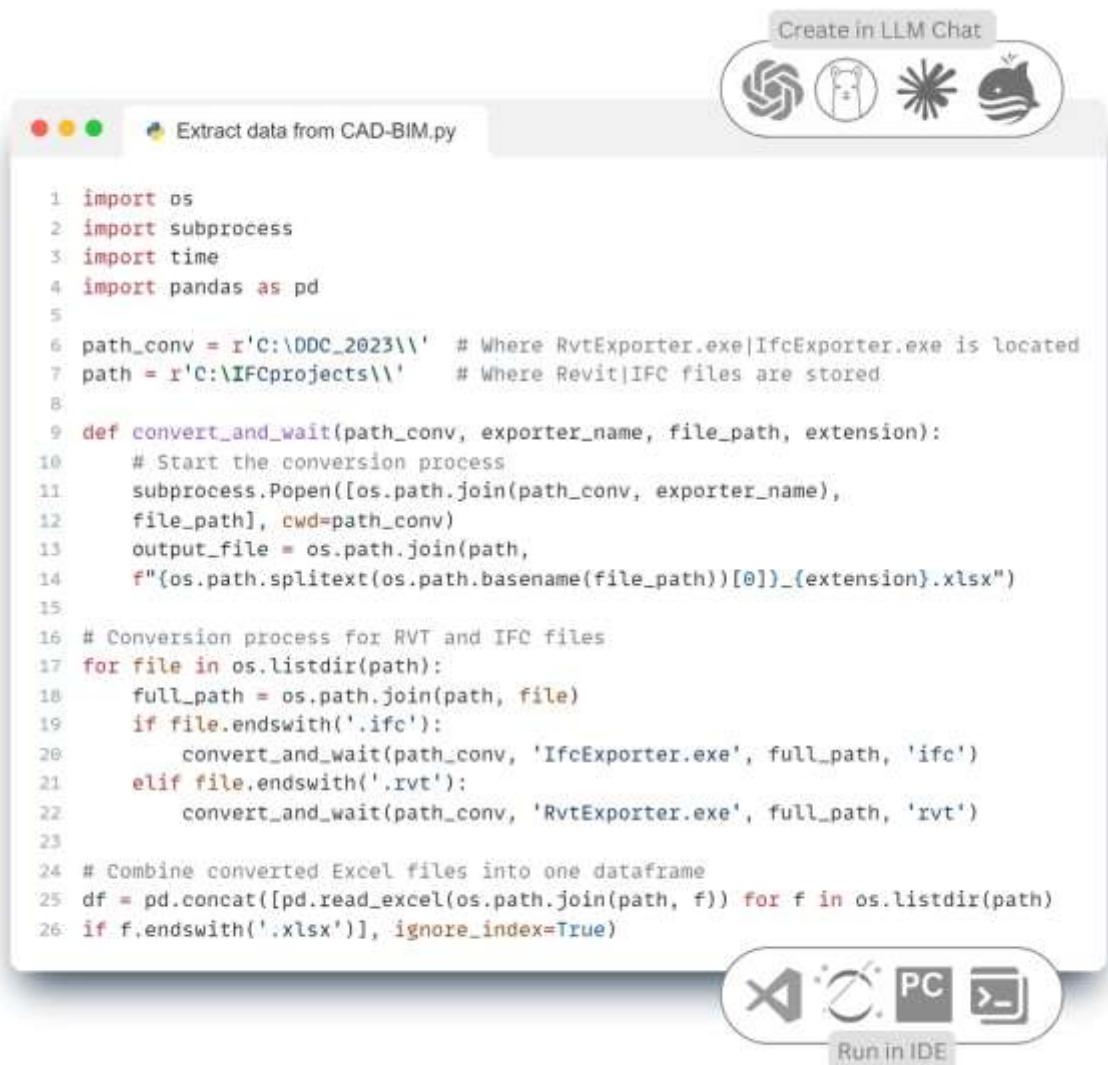


Fig. 7.2-4 Konverter CAD-data (BIM) til én stor dataramme som skal inneholde alle prosjektdelene.

Som et illustrerende eksempel bruker vi datainlastingstrinnet Extract og innhenter en tabell over alle CAD-(BIM-) prosjekter (fig. 7.2-4) ved hjelp av reverse engineering-aktiverte konverteringsprogrammer [138] for RVT- og IFC-formater for å hente strukturerte tabeller fra alle prosjekter og kombinere dem i en stor DataFrame-tabell.



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\' # Where RvtExporter.exe\IfcExporter.exe is located
7 path = r'C:\IFCprojects\' # Where Revit\IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10    # Start the conversion process
11    subprocess.Popen([os.path.join(path_conv, exporter_name),
12                      file_path], cwd=path_conv)
13    output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16    # Conversion process for RVT and IFC files
17    for file in os.listdir(path):
18        full_path = os.path.join(path, file)
19        if file.endswith('.ifc'):
20            convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21        elif file.endswith('.rvt'):
22            convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24    # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Fig. 7.2-5 Konvertering ved hjelp av Python kode og SDK reverse engineering-verktøy for RVT og IFC filer til en stor strukturert (df) DataFrame.

Pandas DataFrame kan laste inn data fra en rekke kilder, inkludert CSV-tekstfiler, Excel-regneark, JSON - og XML - filer, stordatalagringsformater som Parquet og HDF5, og fra MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle og andre databaser. I tillegg støtter Pandas innlasting av data fra API-er, nettsider, skytjenester og lagringssystemer som Google BigQuery, Amazon Redshift og Snowflake.

- ☞ Hvis du vil skrive kode for å koble til og samle inn informasjon fra databaser, kan du sende en lignende tekstforespørrelse til LLM chatterom (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN eller andre):

Vennligst skriv et eksempel på tilkobling til MySQL og konvertering av data til ↵

💡 LLMs svar:

```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7     user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

Run in IDE

Figur 7.2-6 Eksempel på tilkobling via Python til en MySQL-database og import av data fra MySQL-databasen til en DataFrame .

Den resulterende koden (fig. 7.2-5, fig. 7.2-6) kan kjøres i en av de populære IDE-ene (integrerte utviklingsmiljøene) vi nevnte ovenfor i frakoblet modus: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære nettbaserte verktøy: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Ved å laste multiformatdataene inn i variabelen "df" (Figur 7.2-5 - rad 25; Figur 7.2-6 - rad 8), konverterte vi dataene til formatet Pandas DataFrame, en av de mest populære strukturene for databehandling, som er en todimensjonal tabell med rader og kolonner. Vi kommer til å snakke mer om andre lagringsformater som brukes i ETL -Pipelines, som Parquet, Apache ORC, JSON, Feather, HDF5, samt moderne datavarehus i kapittelet "Datalagring og -håndtering i byggebransjen" (figur 8.1-2).

Etter datauttrekket og struktureringen (Extract) dannes en enkelt informasjonsmatrise (fig. 7.2-5, fig. 7.2-6), som er klar for videre behandling. Før dataene lastes inn i målsystemene eller brukes til analyse, er det imidlertid nødvendig å sikre at kvaliteten, integriteten og samsvaret med de spesifiserte kravene er ivaretatt. Det er på dette stadiet datatransformasjonen (Transform) finner sted - et viktig skritt for å sikre påliteligheten til senere konklusjoner og beslutninger.

ETL Transform: anvendelse av validerings- og transformasjonsregler

I trinnet Transform blir dataene behandlet og transformert. Denne prosessen kan omfatte korrekthetskontroll, normalisering, utfylling av manglende verdier og validering hjelpt av automatiserte

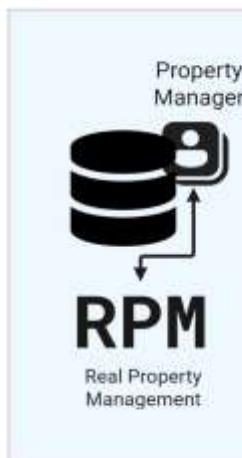
verktøy.

Ifølge PwC-studien "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] beveger moderne revisjonsselskaper seg bort fra tilfeldig dataverifikasi og over til å analysere enorme mengder informasjon ved hjelp av automatiserte verktøy. Denne tilnærmingen gjør det ikke bare mulig å identifisere avvik i rapporteringen, men også å gi anbefalinger om hvordan forretningsprosessene kan optimaliseres.

I bygg- og anleggsbransjen kan lignende metoder brukes til for eksempel automatisk validering av prosjekteringsdata, kvalitetskontroll og evaluering av entreprenørens prestasjoner. Et verktøy for å automatisere og øke hastigheten på databehandlingen er bruk av regulære uttrykk (RegEx) i datatransformasjonsfasen (Transform) i ETL-prosessen. Med RegEx kan du effektivt validere datastrenger, oppdage inkonsekvenser og sikre informasjonens integritet med minimalt ressursforbruk. Flere detaljer om RegEx (fig. 4.4-7) ble diskutert i kapittelet "Oversette krav til strukturert form".

La oss se på et praktisk eksempel: I systemet for forvaltning av eiendomsobjekter (RPM) setter forvalteren kravene for nøkkellattributter for objekter (fig. 7.2-7). På transformasjonsstadiet må følgende parametere valideres:

- verifikasiing av formatet på objektidentifikatoren (attributtet "ID")
- kontroll av verdier for erstatningsgarantiperiode (attributtet "Garantiperiode")
- verifikasiing av syklusen for utskifting av elementer (attributtet "Vedlikeholdskrav")



Property Manager:
Long-term Management

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnistørløp

Figur 7.2-7 Verifikasiing av kvalitet begynner med å sette krav til attributter og grenseverdier for disse.

For å angi grenseverdier for parametervalidering kan vi for eksempel anta at vi vet fra at gyldige verdier for "ID"-attributtet bare kan inneholde strengverdiene "W-NEW", "W-OLD1" eller "D-122" eller lignende verdier der det første tegnet er en bokstav etterfulgt av en bindestrek og deretter tre alfabetiske tegn "NEW", "OLD" eller et tresifret tall (figur 7.2-7). Følgende regulære uttrykk (RegEx) kan brukes til å validere disse identifikatorene:

```
^W-NEW$| ^W-OLD[0-9]+$| ^D-1[0-9]{2}$
```

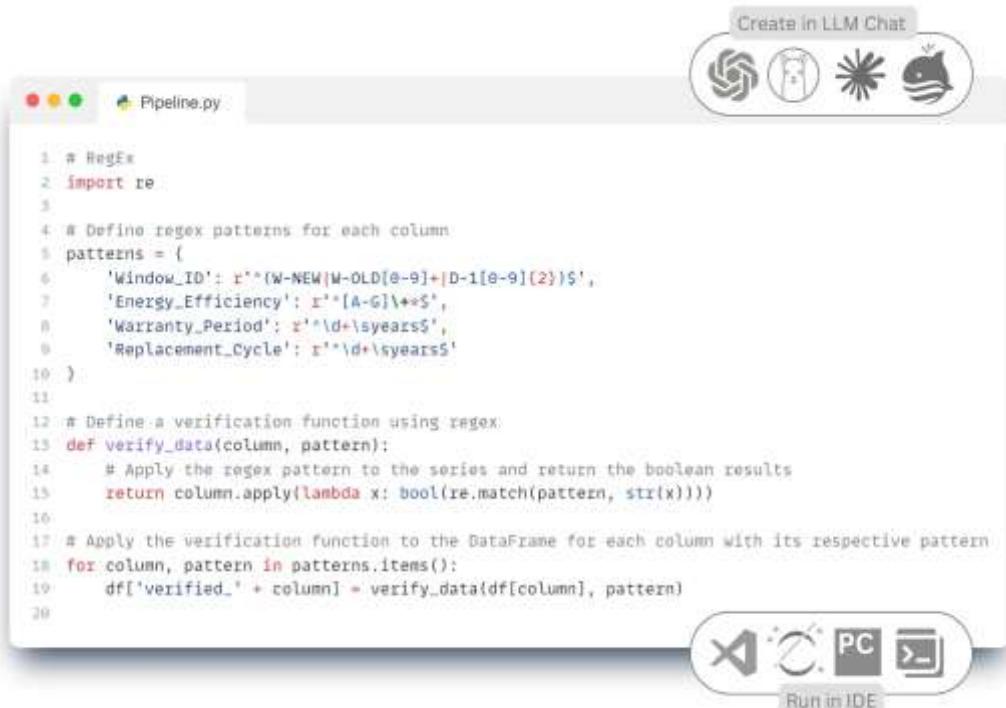
Denne malen sørger for at alle identifikatorer i dataene samsvarer med de angitte kriteriene. Hvis en verdi ikke består kontrollen, logger systemet en feilmelding. Hvis du vil lage Python -kode for å transformere dataene og bruke de resulterende dataene til å lage en resultattabell, formulerer du ganske enkelt en spørring i LLM -chatten.

 Tekstforespørsel til LLM:

Skriv kode for å validere kolonnene i DataFrame ved hjelp av regulære uttrykk, som kontrollerer identifikatorer i formatet 'W-NEW' eller 'W-OLD' via RegEx, energieffektivitet med bokstavene 'A' til 'G', garantiperiode og utskiftingssyklus med numeriske verdier i år



 LLMs svar :



```

1 # Regex
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^[W-NEW|W-OLD]{0-9}+[D-1]{0-9}(2?)$',
7     'Energy_Efficiency': r'^[A-G]\w*$',
8     'Warranty_Period': r'^\d+years$',
9     'Replacement_Cycle': r'^\d+\syears$'
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20

```

Create in LLM Chat



Run in IDE

Figur 7.2-8 Code automatiserer valideringsprosessen ved å bruke RegEx-maler på datarammeparameterkolonnene .

Den automatisk genererte Python -koden ovenfor (figur 7.2-8) bruker "re"-biblioteket (RegEx regular expressions) til å definere en funksjon som sjekker hvert attributt i et dataelement i en DataFrame. For hver kolonne (attributt) som er spesifisert, bruker funksjonen et RegEx-mønster for å kontrollere at hver post er i samsvar med det forventede formatet, og legger til resultatene som nye verdier (False/True) i den nye attributt-kolonnen i DataFrame.

En slik automatisert validering sikrer formell datakompatibilitet og kan brukes som en del av et kvalitetskontrollsysten i transformasjonsfasen.

Etter at trinnet Transform og kvalitetskontrollen er fullført, er dataene klare til å lastes opp til målsystemene. Transformerte og validerte data kan lastes opp til CSV, JSON, Excel, databaser og andre formater for videre bruk. Avhengig av oppgaven kan resultatene også presenteres i rapporter, grafer eller analytiske dashbord.

ETL Load: Visualiser resultater i diagrammer og grafer

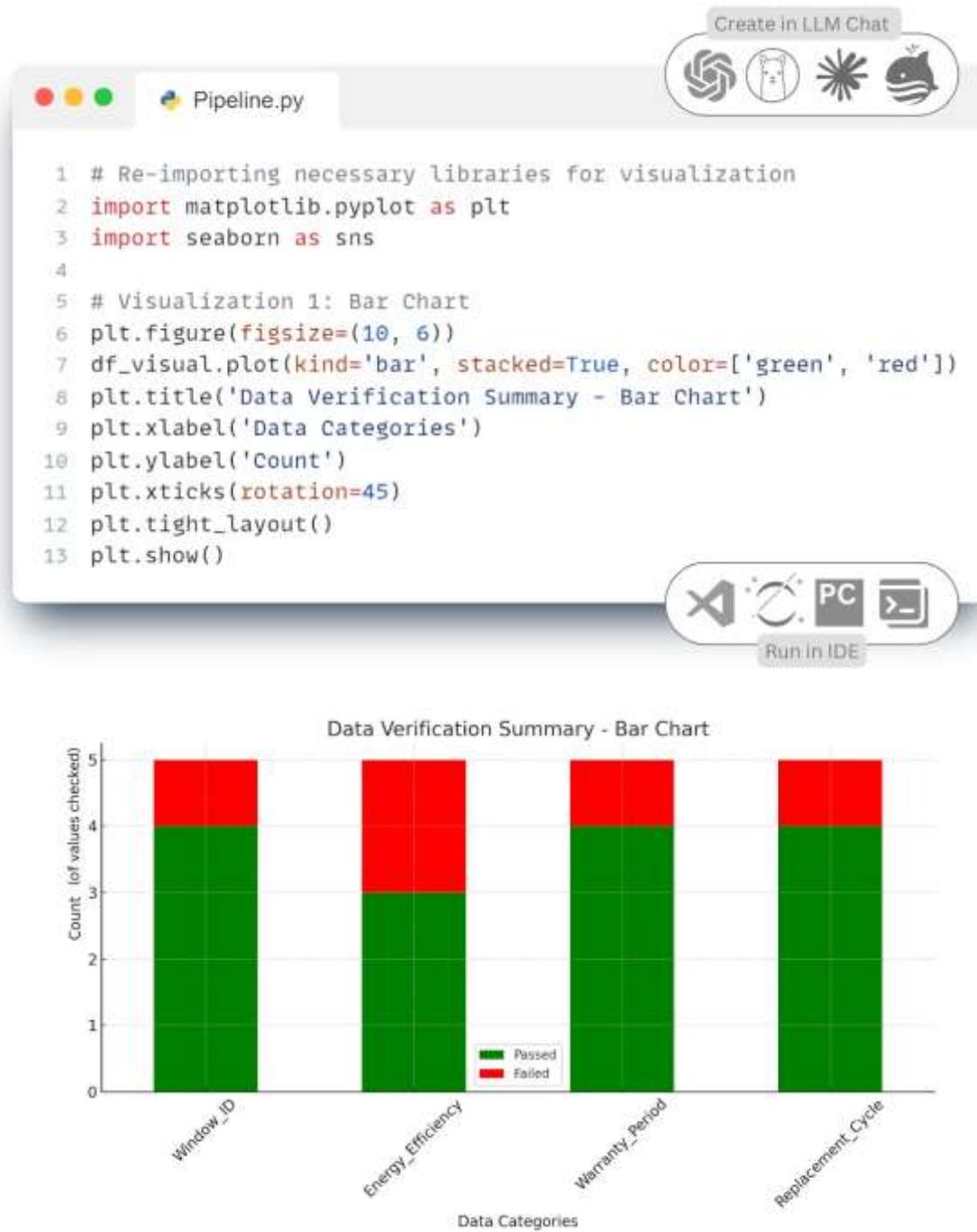
Etter at Transform er fullført, når dataene er strukturert og verifisert, er det siste steget Load, der dataene både kan lastes inn i målsystem og visualiseres for analyse. Visuell presentasjon av dataene gjør det mulig å raskt identifisere avvik, analysere fordelinger og kommunisere viktige konklusjoner til alle prosjektdeltakere, også dem uten teknisk bakgrunn.

I stedet for å presentere informasjon i form av tabeller og tall, kan vi bruke infografikk, grafer og instrumentpaneler (dashboards). Et av de vanligste og mest fleksible verktøyene for visualisering av strukturerte data i Python er Matplotlib-biblioteket (figur 7.2-9, figur 7.2-10). Det gjør det mulig å lage statiske, animerte og interaktive diagrammer, og støtter et bredt spekter av diagramtyper.

- For å visualisere resultatene av attributtkontrollen fra RPM-systemet (figur 7.2-7), kan du bruke følgende spørring til språkmodellen:

Skriv kode for å visualisere DataFrame-dataene ovenfor (figur 7.2-7), med et histogram for resultatene for å vise hyppigheten av feil i attributtet ↴

- ❷ LLM-svar i kodeform og klar visualisering direkte i LLM-chatrommet av resultatene av kodekjøringen :



The screenshot shows a code editor window titled "Pipeline.py" containing the following Python code:

```

1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()

```

Below the code editor are several icons: "Create in LLM Chat" (with a brain icon), a cat icon, a starburst icon, a whale icon, a "Run in IDE" button (with a play icon), and a "PC" icon.

At the bottom is a stacked bar chart titled "Data Verification Summary - Bar Chart". The Y-axis is labeled "Count (of values checked)" and ranges from 0 to 5. The X-axis is labeled "Data Categories" and includes "Window_ID", "Energy_Efficiency", "Warranty_Period", and "Replacement_Cycle". The legend indicates "Passed" (green) and "Failed" (red). The chart shows that for each category, there are approximately 4 Passed values and 1 Failed value.

Data Category	Passed	Failed
Window_ID	4	1
Energy_Efficiency	3	2
Warranty_Period	4	1
Replacement_Cycle	3	1

Fig. 7.2-9 Visualisering av resultatene av Transform -trinnet for kontroll av attributtverdier fra RPM-systemet (Fig. 7.2-7) som et histogram i Load -trinnet .

- Det finnes mange gratis visualiseringsbiblioteker med åpen kildekode som gjør det mulig å presentere strukturerte data i ulike formater. La oss fortsette å visualisere resultatene med en annen type graf med følgende ledetekst i chatten:

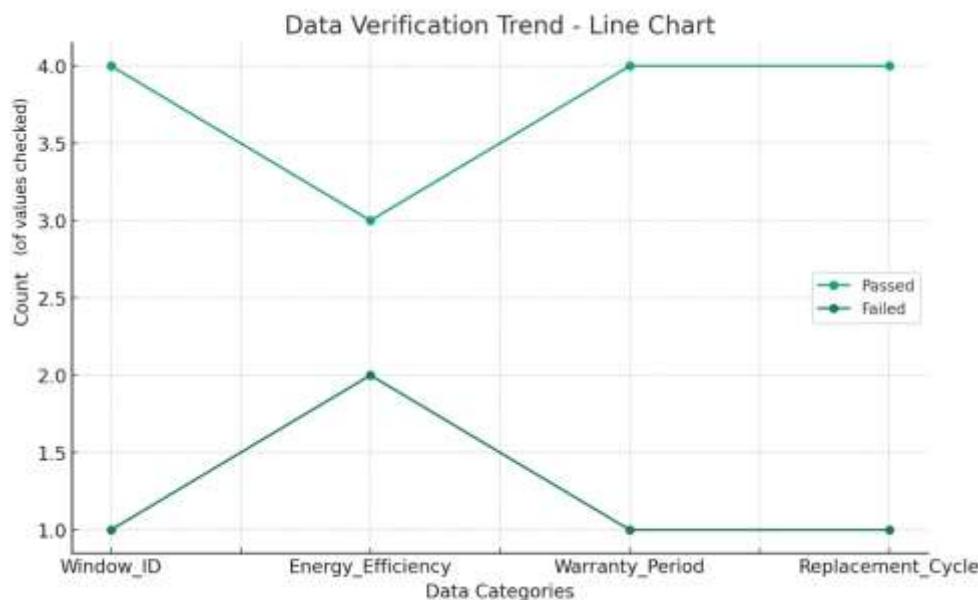
Plott de samme dataene som en graf med linjer ↴

LLMs svar :

```

1 # Visualization 3: Line Chart
2 df_visual.plot(kind='line', marker='o', figsize=(10, 6))
3 plt.title('Data Verification Trend - Line Chart')
4 plt.xlabel('Data Categories')
5 plt.ylabel('Count')
6 plt.grid(True)
7 plt.show()

```



Figur 7.2-10 Visualisering av valideringsdataene (Figur 7.2-8) som et linjediagram ved hjelp av Matplotlib-biblioteket .

Det finnes mange gratis visualiseringsbiblioteker med åpen kildekode, som f.eks:

- Seaborn - for statistiske grafer (Fig. 7.2-11)
- Plotly - for interaktive webvisualiseringer (fig. 7.2-12, fig. 7.1-6)
- Altair - for deklarativ visualisering
- Dash eller Streamlit - for å lage komplette dashboard

Det er ikke nødvendig å ha kjennskap til spesifikke biblioteker for visualisering - moderne verktøy, blant annet LLM, lar deg automatisk generere kode for grafer og hele applikasjoner basert på oppgavebeskrivelsen.

Valget av verktøy avhenger av formålet med prosjektet: om det er en rapport, en presentasjon eller et dashboard på nettet. For eksempel er open source-biblioteket Seaborn spesielt godt egnet til å arbeide med kategoriske data, slik at man kan identifisere mønstre og trender.

- ▣ For å se Seaborn-biblioteket i drift, kan du enten be LLM om å bruke det ønskede biblioteket direkte eller sende en tilsvarende tekstforespørsel til LLM på:



- ▣ LLMs svar i form av kode og en ferdig graf, hvor plottekoden nå kan kopieres inn i IDE, og selve grafen kan kopieres eller lagres for å limes inn i et dokument:

```

1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

```



Figur 7.2-11 Visualisering av resultatene av valideringen (Figur 7.2-8) av dataene ved hjelp av Seaborn-biblioteket.

For de som foretrekker en interaktiv tilnærming, finnes det verktøy som gjør det mulig å lage dynamiske diagrammer og paneler med mulighet for interaksjon. Plotly-biblioteket (figur 7.1-6, figur 7.2-12) gjør det mulig å lage svært interaktive diagrammer og paneler som kan integreres i nettsider og gi brukeren mulighet til å interagere med dataene i sanntid.

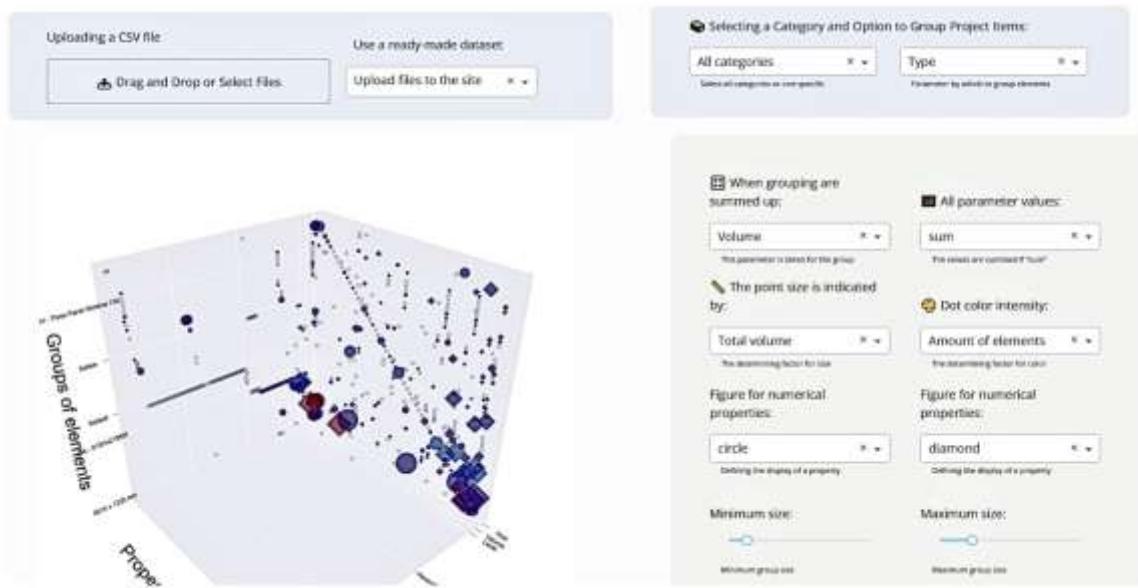


Fig. 7.2-12 Interaktiv 3D -visualisering av elementattributter fra et CAD- (BIM-) prosjekt ved hjelp av Plotly-biblioteket.

Spesialbibliotekene Bokeh, Dash og Streamlit med åpen kildekode gjør det enkelt å presentere data uten å måtte ha inngående kjennskap til webutvikling. Bokeh eigner seg for komplekse interaktive grafer, Dash brukes til å bygge fullverdige analytiske dashbord, og med Streamlit kan du raskt lage webapplikasjoner for dataanalyse.

Med slike visualiseringsverktøy kan utviklere og analyticere effektivt formidle resultater til kolleger og interessenter, slik at de kan samhandle intuitivt med dataene og forenkle beslutningsprosessen.

ETL Load: Automatisk opprettelse av PDF-dokumenter

På datalastingsstadiet kan du ikke bare visualisere dataene, laste dem opp til tabeller eller databaser, men også automatisk generere rapporter, inkludert nødvendige grafer, diagrammer og viktige analytiske indikatorer som skal mottas av lederen eller spesialisten som venter på resultatene av revisjonen. Automatiserte rapporter kan inneholde både kommentarer og tekstlig tolkning av data og visualiseringselementer - tabeller og grafer.

- For å lage en PDF -rapport med et histogram (fig. 7.2-9) og en beskrivelse av analysen basert på valideringen vi utførte i de foregående kapitlene, er det tilstrekkelig å formulere en forespørsel i en dialogfortsettelse med LLM, f.eks.:

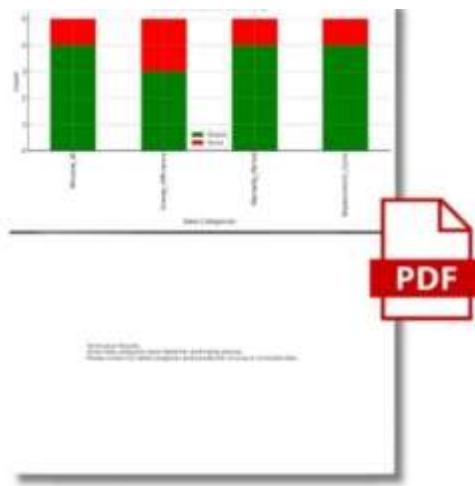
Skriv kode for å opprette en PDF -fil med et histogram og en beskrivelse av resultatene av datavalideringen ovenfor (i chatten), og skriv en advarselstekst om at noen kategorier ikke er validert, og at manglende må fylles ut ↵

- LLMs svar som kode og ferdig PDF med resultater:

```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df.pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "Please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



Figur 7.2-13 Den automatiserte koden oppretter et PDF -dokument som inneholder et histogram med testdata og tekst med testresultater.

En automatisk skrevet løsning på bare 20 kodelinjer ved hjelp av LLM skaper umiddelbart det ønskede PDF (eller DOC)-dokumentet med visualisering i form av et attributhistogram (fig. 7.2-13) som viser antall data som har bestått og ikke bestått valideringen, og med tillegg av en tekstblokk som oppsummerer resultatene og anbefalinger videre tiltak.

Automatisert dokumentgenerering er et sentralt element i Load stage, spesielt i et prosjektmiljø der rask rapportering og nøyaktighet er avgjørende.

ETL Load: automatisk generering av dokumenter fra FPDF

Automatisering av rapportering på ETL-stadiet Load er et viktig trinn i databehandlingen, spesielt når resultatene av analysene skal presenteres i et format som er enkelt å kommunisere og forstå. I byggebransjen er dette ofte relevant for fremdriftsrapporter, prosjektdatastatistikk, kvalitetssikringsrapporter eller økonomisk dokumentasjon.

Et av de mest praktiske verktøyene for slike oppgaver er open source-biblioteket, FPDF, som er tilgjengelig for både Python og PHP.

FPDF-biblioteket er en fleksibel måte å generere dokumenter på ved hjelp av kode, slik at du kan legge til overskrifter, tekst, tabeller og bilder. Ved å bruke kode i stedet for manuell redigering reduseres antall feil, og prosessen med å utarbeide rapporter i PDF -format går raskere.

Et av de viktigste trinnene i opprettelsen av et PDF -dokument er å legge til overskrifter og hovedteksten i form av kommentarer eller beskrivelser. Når du lager en rapport, er det imidlertid ikke bare viktig å legge til tekst, men også å strukturere den på riktig måte. Overskrifter, innrykk, linjeavstand - alt dette påvirker lesbarheten til dokumentet. Ved hjelp av FPDF kan du angi formateringsparametere, kontrollere plasseringen av elementer og tilpasse dokumentets stil.

FPDF er i prinsippet svært lik HTML. De som allerede er kjent med HTML, kan enkelt generere PDF -

dokumenter av enhver kompleksitet ved hjelp av FPDF, ettersom kodestrukturen er svært lik HTML-markup: overskrifter, tekst, bilder og tabeller legges til på samme måte. De som ikke er kjent med HTML, kan bruke LLM, som umiddelbart hjelper deg med å komponere koden for å generere ønsket dokumentlayout.

- Følgende eksempel viser hvordan du genererer en rapport med overskrift og brødtekst. Ved å kjøre denne koden i en IDE med Python-støtte opprettes en PDF -fil som inneholder ønsket overskrift og tekst:

```
from fpdf import FPDF      # Importer FPDF-biblioteket
pdf = FPDF()    # Opprett PDF -dokument
pdf.add_page()  # Legg til en side

pdf.set_font("Arial", style='B', size=16) # Sett font: Arial, fet, størrelse 16
pdf.cell(200, 10, "Prosjektrapport", ln=True, align='C')  # Opprett en overskrift og sentrer den
pdf.set_font("Arial", size=12) # Endre skriftypen til vanlig Arial, størrelse 12
pdf.multi_cell(0, 10, "Dette dokumentet inneholder data om resultatene av verifiseringen av prosjektfilen . . .") # Legg til tekst på flere linjer
pdf.output(r "C:\reports\report.pdf") # Lagre PDF -fil
```



Figur 7.2-14 Med noen få linjer Python-kode kan vi automatisk generere PDF-filen tekstdokumentet vi trenger.

Når man utarbeider rapporter, er det viktig å ta hensyn til at dataene som dokumentet danner fra, sjeldent er statiske. Overskrifter, tekstblokker (fig. 7.2-14) danner ofte dynamisk, og mottar verdier i transformasjonsfasen i ETL-prosessen.

Ved hjelp av koden kan du opprette dokumenter som inneholder oppdatert informasjon: prosjektnavn, dato for generering av rapporten, samt informasjon om deltakere eller nåværende status. Ved å bruke variabler i koden kan du automatisk sette inn disse dataene på de ønskede stedene i rapporten, slik at du slipper å redigere dem manuelt før du sender dem.

I tillegg til enkel tekst og overskrifter har tabeller en spesiell plass i prosjektdokumentasjonen. Nesten alle dokumenter inneholder strukturerte data: fra objektbeskrivelser til inspeksjonsresultater. Automatisk generering av tabeller basert på data fra Transform -fasen gjør det ikke bare mulig å fremskynde prosessen med å utarbeide dokumenter, men også å minimere feil ved overføring av informasjon. FPDF gjør det mulig å sette inn tabeller i PDF -filer (som tekst eller bilder), og angi cellekanter, kolonnestørrelser og skriftypen (fig. 7.2-15). Dette er spesielt praktisk når man arbeider med dynamiske data, der antall rader og kolonner

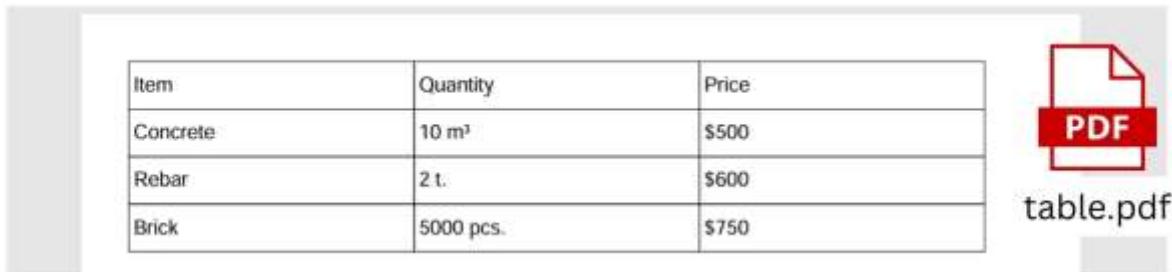
kan variere avhengig av dokumentets oppgaver.

- Følgende eksempel viser hvordan du kan automatisere opprettelsen av tabeller, f.eks. med materiallister, estimerer eller parameterresultater:

```
data = [
    [ "Item", "Quantity", "Price" ],   # Kolonneoverskrifter
    [ "Beton", "10 m³", "$ 500." ],   # Data for første rad
    [ "Armeringsjern", "2 tonn", "$ 600." ],   # Data for andre rad.
    [ "Murstein", "5 000 stykker", "$ 750." ],   # Data på linje 3.
]

pdf = FPDF () # Opprett PDF -document
pdf.add_page() # Legg til en side
pdf.set_font("Arial", size=12) # Angi skriftype

for row in data: # Søk i tabellradene
    for item in row: # Gå gjennom cellene i raden
        pdf.cell(60, 10, item, border=1) # Opprett en celle med en ramme, bredde 60 og høyde 10
    pdf.ln() # Gå til neste linje
pdf.output(r "C:\reports\table.pdf") # Lagre PDF -file
```



Item	Quantity	Price
Concrete	10 m³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

Figur 7.2-15 Du kan automatisk generere ikke bare tekst, men også tabellinformasjon fra Transform-trinnet i PDF.

I reelle rapporteringsscenarier er tabeller vanligvis dynamisk generert informasjon som innhentes på datatransformasjonsstadiet. I eksemplet (fig. 7.2-15) er tabellen satt inn i PDF -dokumentet i en statisk form: dataene for eksemplet ble plassert i dataordboken (første linje i koden), i virkelige forhold fylles slike datavariabler inn automatisk etter f.eks. gruppering av datarammen.

I praksis bygges slike tabeller ofte på grunnlag av strukturerte data som kommer fra ulike dynamiske kilder: databaser, Excel -filer, API -grensesnitt eller resultater av analytiske beregninger. Oftest aggregeres, gruppertes eller filtreres dataene i Transform (ETL)-fasen - og først deretter omdannes de til totaler i form av grafer eller todimensjonale tabeller som vises i rapporter. Det betyr at tabellinnholdet kan endres avhengig av valgte parametere, analyseperiode, prosjektfiltre eller brukerinnstillinger.

Bruken av dynamiske datarammer og datasett i Transform -fasen gjør rapporteringsprosessen i Load -fasen så fleksibel, skalerbar og lett å gjenta som mulig uten behov for manuell inngripen.

I tillegg til tabeller og tekst støtter FPDF også grafer av tabelldata, slik at du kan legge inn bilder generert med Matplotlib eller andre visualiseringsbiblioteker som vi har omtalt ovenfor, i rapporten. Alle grafer, diagrammer og diagrammer kan legges til i dokumentet ved hjelp av koden.

- Ved hjelp av Python-biblioteket FPDF kan vi legge til en graf som er forhåndsgenerert med Matplotlib. i PDF-dokumentet:

```
import matplotlib.pyplot as plt # Importer Matplotlib for å lage grafer

fig, ax = plt.subplots() # Opprett figuren og aksene i grafen
categories = ["Betong", "Armeringsjern", "Murstein"] # Kategorinavn
values = [50000, 60000, 75000] # Kategoriens verdier
ax.bar(categories, values) # Lag et søylediagram
plt.ylabel("Verdi,$.") # Signer Y-aksen
plt.title("Kostnadsfordeling") # Legg til en tittel
plt.savefig(r "C:\reports\chart\chart\chart.png") # Lagre diagrammet som et bilde

pdf = FPDF () # Create PDF -document
pdf.add_page() # Legg til en side
pdf.set_font("Arial", size=12) # Angi skrifftype
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # Legg til en topptekst

pdf.image(r "C:\reports\chart\chart\chart\chart.png", x=10, y=30, w=100) # Sett
bildet inn i PDF-filen (x, y - koordinater, w - bredde)
pdf.output(r "C:\reports\chart_report.pdf") # Lagre PDF-filen
```

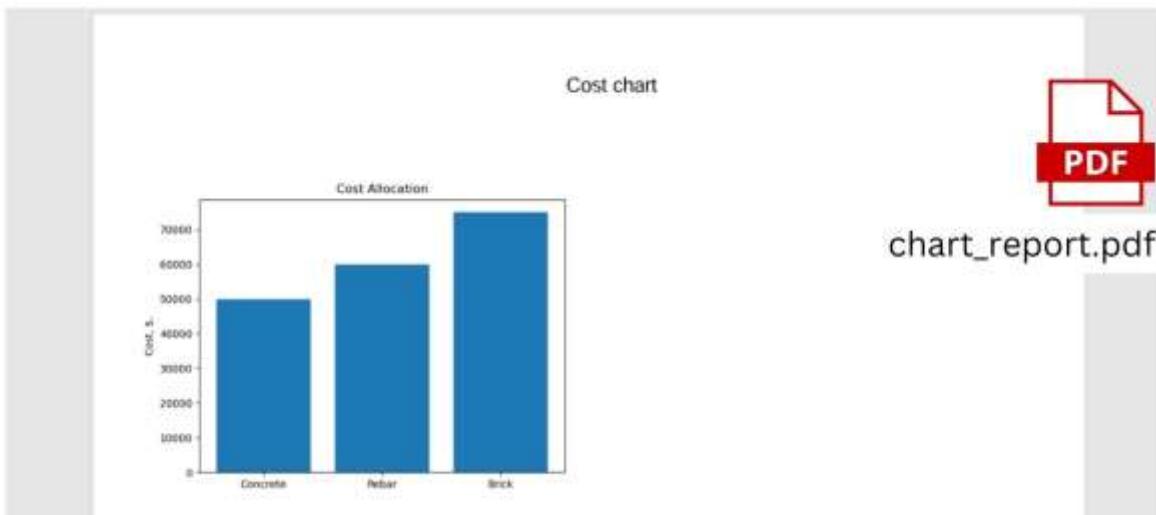


Fig. 7.2-16et titalls kodelinjer kan du generere en graf, lagre den og lime den inn i et PDF-dokument.

Med FPDF blir prosessen med dokumentforberedelse og -logikk gjennomsiktig, rask og praktisk. Malene som er innebygd i koden, gjør det mulig å generere dokumenter med oppdaterte data, noe som eliminerer behovet for manuell utfylling.

Ved hjelp av ETL-automatisering - i stedet for tidkrevende manuell rapportering kan fagfolk fokusere på å analysere data og ta beslutninger, i stedet for å velge riktig verktøy for å jobbe med en bestemt datasilo med et oversiktlig brukergrensesnitt.

FPDF-biblioteket er dermed et fleksibelt verktøy for automatisert oppretting av dokumenter av enhver kompleksitet - fra korte tekniske rapporter til komplekse analytiske sammendrag med tabeller og diagrammer, noe som ikke bare gjør det mulig å øke hastigheten på dokumentflyten, men også redusere sannsynligheten for feil forbundet med manuell dataregistrering og formatering betraktelig.

ETL Load: Rapportering og innlasting til andre systemer

I Load-stadiet ble resultatene generert i form av tabeller, grafer og endelige PDF-rapporter utarbeidet i samsvar med de etablerte kravene. Det er også mulig å eksportere disse dataene til maskinlesbare formater (f.eks. CSV), noe som er nødvendig for integrering med eksterne systemer som ERP, CAFM, CPM, BI-plattformer og andre bedrifts- eller bransjeløsninger. I tillegg til CSV kan data lastes opp til XLSX, JSON, XML eller direkte til databaser som støtter automatisk informasjonsutveksling.

- ☞ For å generere riktig kode for å automatisere innlastingstrinnet, spør du ganske enkelt LLM - grensesnittet, for eksempel: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude eller QWEN:

Skriv kode for å generere en rapport over datavalideringsresultater i DataFrame, der kolonner med prefikset 'verified_' telles, omdøpes til 'Passed' og 'Failed', manglende verdier erstattes med 0, og deretter eksporteres bare de radene som består alle valideringer til en CSV -fil.

■ LLMs svar:



```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a csv file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

The screenshot shows a Jupyter Notebook cell with the code provided above. The cell has a light gray background and contains several colored lines of Python code. At the bottom of the cell, there are four icons: a red X, a green checkmark, a blue circular arrow, and a black square with a white arrow pointing right. Below these icons, the text "Run in IDE" is visible.

Figur 7.2-17 Validerte data som er innhentet i transformasjonstrinnet fra den endelige datarammen, eksporteres til en CSV -fil for integrering med andre systemer.

Koden ovenfor (fig. 7.2-17) implementerer den siste fasen av ETL -prosessen - Load, der de verifiserte dataene lagres i CSV-format, som er kompatibelt med de fleste eksterne systemer og databaser. Dermed har vi fullført hele ETL -prosessen, inkludert utvinning, transformasjon, visualisering, dokumentasjon og eksport av data til de systemene og formatene vi trenger, noe som sikrer reproducertbarhet, gjennomsiktighet og automatisering av arbeidet med informasjon.

ETL pipeline kan brukes både til behandling av enkelprosjekter og til storskalaapplikasjoner - når man analyserer hundrevis eller tusenvis av innkommende data i form av dokumenter, bilder, skanninger, CAD prosjekter, punktskyer, PDF filer eller andre kilder som kommer fra distribuerte systemer. Muligheten til å automatisere prosessen fullt ut gjør at ETL ikke bare er et teknisk prosesseringsverktøy, men selve grunnlaget for en digital infrastruktur for bygg- og anleggsinformasjon.

ETL med LLM: Visualisering av -data fra PDF -dokumenter

Det er på tide å gå videre til å bygge en fullverdig ETL-prosess som dekker alle viktige trinn i datahåndteringen i ett og samme scenario - uttrekk, transformasjon og innlasting. La oss bygge en automatisert ETL-pipeline som lar deg behandle PDF-dokumenter uten manuelt arbeid - trekke ut data fra dokumenter, visualisere, analysere og overføre til andre systemer.

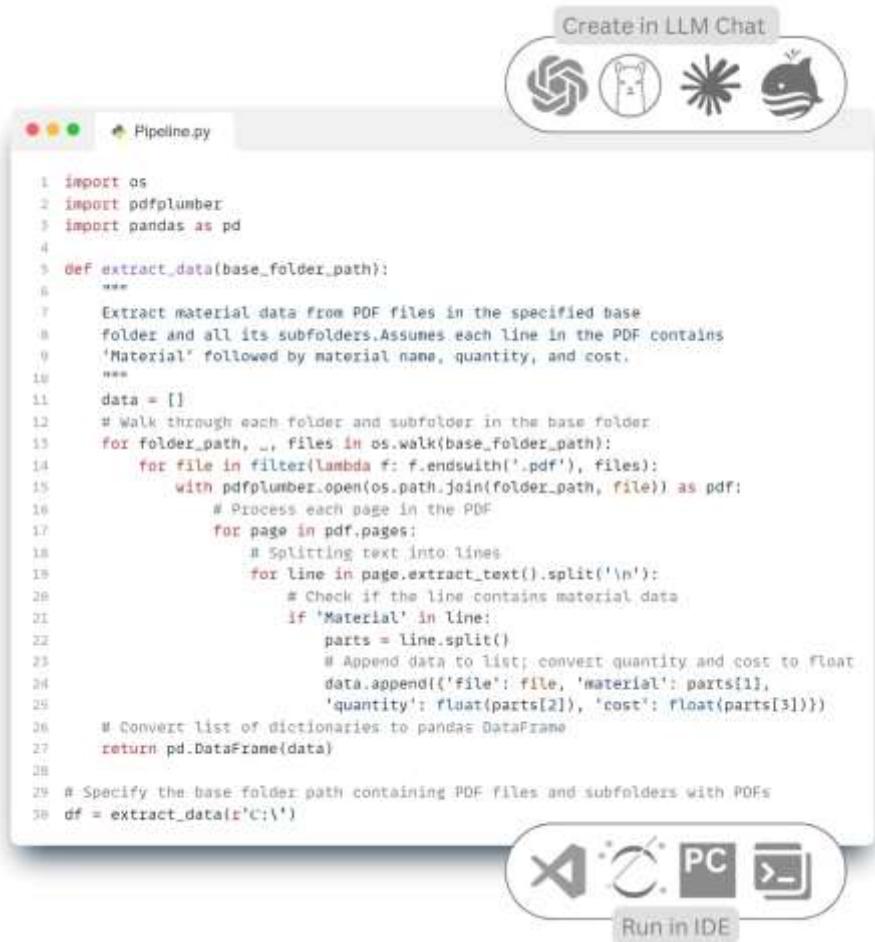
ETL-prosessen i vårt eksempel vil bli beskrevet gjennom prompter, som må forklare språkmodellen (LLM) alle ETL-prosessene med en beskrivelse av sluttresultatet som skal oppnås. I dette tilfellet er oppgaven å

finne alle PDF-filer i den angitte mappen og dens undermapper, trekke ut relevant informasjon fra dem - f.eks. materialnavn, mengder og kostnader - og presentere resultatet som en strukturert tabell (DataFrame) for videre analyse

- 💡 Første tekstspørring i LLM for å automatisk trekke ut data fra flere PDF -dokumenter og opprette en dataramme med data for uttrekkstrinnet:

Skriv kode for å hente ut materialinformasjon fra PDF-filene i en gitt mappe og dens undermapper. Dataene i PDF-filen inkluderer materialnavn, antall og kostnad. Resultatet skal lagres i en DataFrame ↗

- 💡 LLMs svar :



```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26    # Convert list of dictionaries to pandas DataFrame
27    return pd.DataFrame(data)
28
29 # Specify the base folder path containing PDF files and subfolders with PDFs
30 df = extract_data('C:\\')

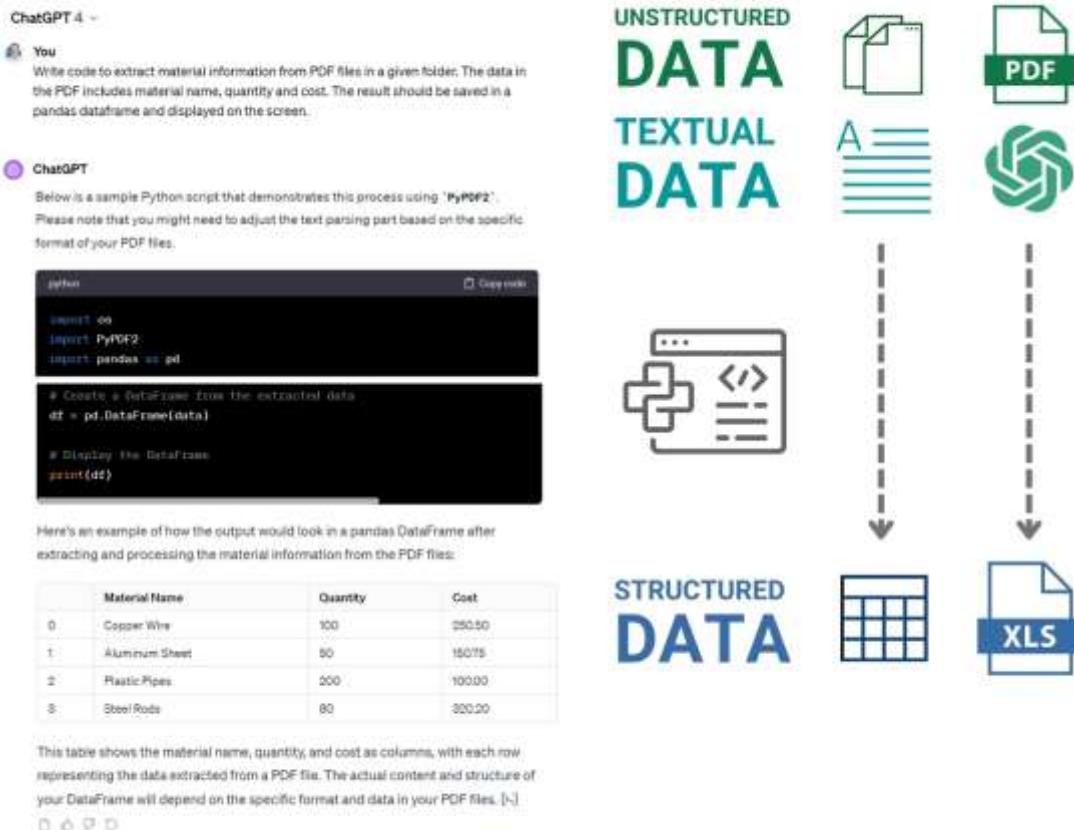
```

Figur 7.2-18 LLM oppretter Python -kode for å hente ut data fra PDF -filer i en bestemt mappe og alle undermappene i denne.

LLM-svaret (figur 7.2-18) er et ferdig Python -skript som automatisk går gjennom alle mapper, åpner PDF -filene som er funnet, trekker ut tekstinformasjon fra dem og konverterer den til en tabell. Den resulterende chat-koden kan kjøres i en av de populære IDE-ene PyCharm, Visual Studio Code (VS Code), Jupyter

Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller populære nettverktøy som Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

For å forenkle innsamlingsprosessen kan vi, i stedet for å kopiere koden fra LLM og bruke koden i IDE, også laste opp et dusin PDF-filer -filer direkte til LLM-chatten (figur 7.2-19) og få en tabell som resultat, uten å måtte se koden eller kjøre den. Resultatet av å kjøre denne koden vil være en tabell med de attributtene vi har valgt.



Figur 7.2-19 Resultatet av kjøring av kode i LLM, som henter ut data fra PDF -filer i en strukturert datarammevisning med utvalgte attributter.

I neste trinn ber vi om en språkmodell for de innhente dataene - for eksempel for å sammenligne kostnader og volum for materialbruk og lage noen eksempler på visualiseringer som kan brukes som grunnlag for videre analyse.

- Be i en fortsatt chat med LLM om å plotte noen grafer fra tabellene som ble produsert i Transform-trinnet (figur 7.2-18):

Visualiser totalkostnaden og mengden av hvert materiale fra DataFrame (figur 7.2-18)

```

1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

Figur 7.2-20 LLM-respons -modeller som Python-kode for å visualisere data fra en dataramme ved hjelp av matplotlib-biblioteket.

LLM genererer og kjører automatisk Python -kode (fig. 7.2-20) ved hjelp av matplotlib-biblioteket. Etter å ha kjørt denne koden får vi grafer over kostnader og materialbruk i byggeprosjekter direkte i chat (fig. 7.2-21), noe som forenkler det analytiske arbeidet betraktelig.

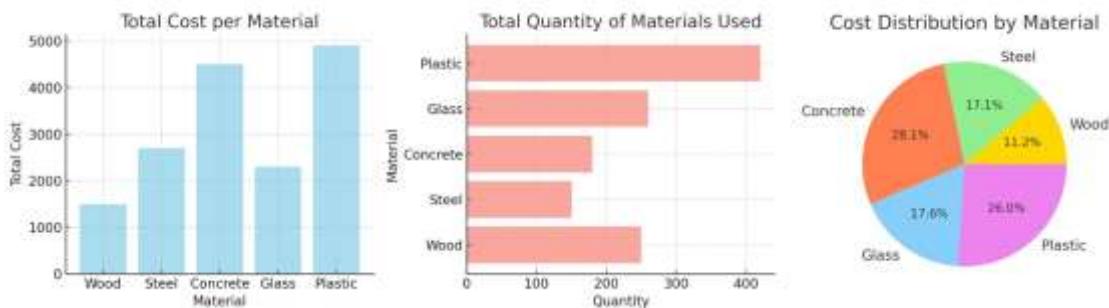
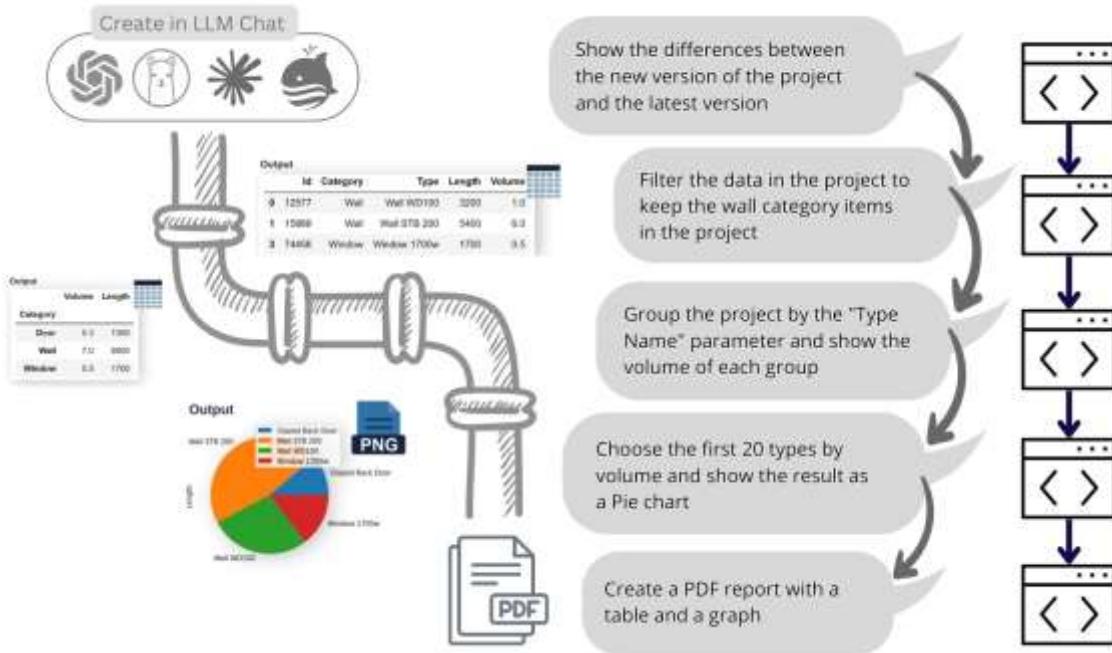


Fig. 7.2-21 Visualisering av LLM-responsen som grafer basert på data som er samlet inn i DataFrame.

Støtte til å utvikle ideer for å skrive ETL -kode, analysere og kjøre kode og visualisere resultater er tilgjengelig gjennom enkle tekstspørninger i LLM, uten at man trenger å lære seg det grunnleggende om programmering. Fremveksten av AI -verktøy som LLM er definitivt i ferd med å endre tilnærmingen til programmering og automatisering av databehandling (figur 7.2-22).

Ifølge PwC-rapporten "What is the real value of artificial intelligence for your business and how can you capitalise on it?" (2017) [139] vil prosessautomatisering og produktivitetsforbedringer være de viktigste drivkraftene bak økonomisk vekst. Og produktivitetsforbedringer forventes å stå for mer enn 55% av all KI-drevet BNP-vekst mellom 2017 og 2030."



Figur 7.2-22 AI LLM bidrar til å generere utkast til kode som kan brukes i fremtidige prosjekter uten at det er behov for en LLM.

Ved hjelp av verktøy som ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok, samt åpne data og programvare med åpen kildekode, kan vi automatisere prosesser som tidligere bare ble gjort med spesialiserte, kostbare og vanskelig vedlikeholdbare modulære proprietære systemer.

I bygge- og anleggsbransjen betyr dette at selskaper som er de første til å implementere automatiserte Pipeline -dataprossesser, vil høste betydelige fordeler, fra forbedret effektivitet i prosjektstyringen til reduserte økonomiske tap og eliminering av fragmenterte applikasjoner og siloformede datavarehus.

Den beskrevne logikken for utførelse av forretningsoppgaver i ETL-prosessenen viktig del av automatiseringen av analyse og databehandlingsprosesser, som er en spesifikk variant av et bredere konsept - rørledninger (Pipelines).

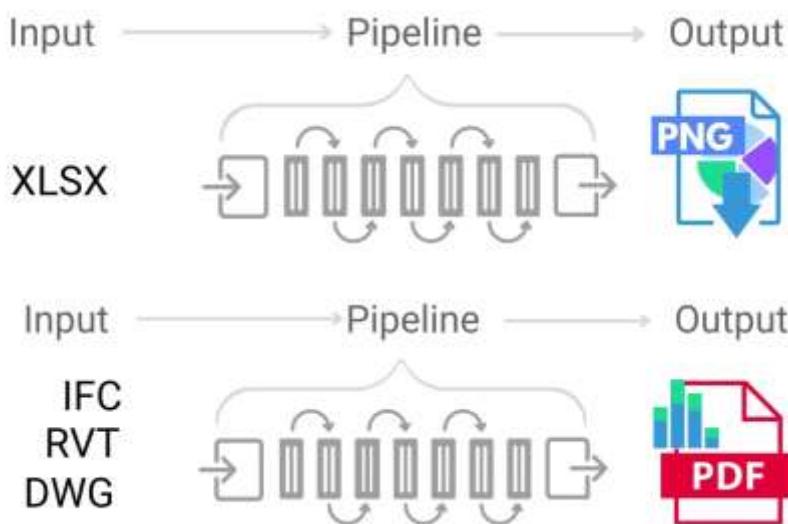
KAPITTEL 7.3.

AUTOMATISK ETL TRANSPORTBÅND (RØRLEDNING)

Pipeline: Automatisk ETL transportbånd data

ETL-prosessen har tradisjonelt blitt brukt til å behandle data i analysesystemer, og dekker både strukturerte og ustukturerte kilder. I dagens digitale miljø brukes imidlertid i økende grad et bredere begrep - Pipeline (transportbånd), som beskriver en hvilken som helst sekvensiell prosesseringskjede der resultatet fra ett trinn blir input til det neste.

Denne tilnærmingen gjelder ikke bare data, men også andre typer automatisering: oppgavebehandling, bygningsrapportering, integrasjon med programvare og digital arbeidsflyt (fig. 7.3-1).



Figur 7.3-1 Pipeline er en prosesseringssekvens der utdataene fra ett trinn blir input til neste trinn.

Bruken av Pipeline er et av hovedelementene i automatisering, spesielt når man arbeider med store mengder heterogene data. Pipeline-arkitekturen gjør det mulig å organisere komplekse behandlingstrinn i et modulært, konsistent og håndterbart format, noe som øker lesbarheten, forenkler vedlikehold av koden og muliggjør inkrementell feilsøking og skalerbar testing.

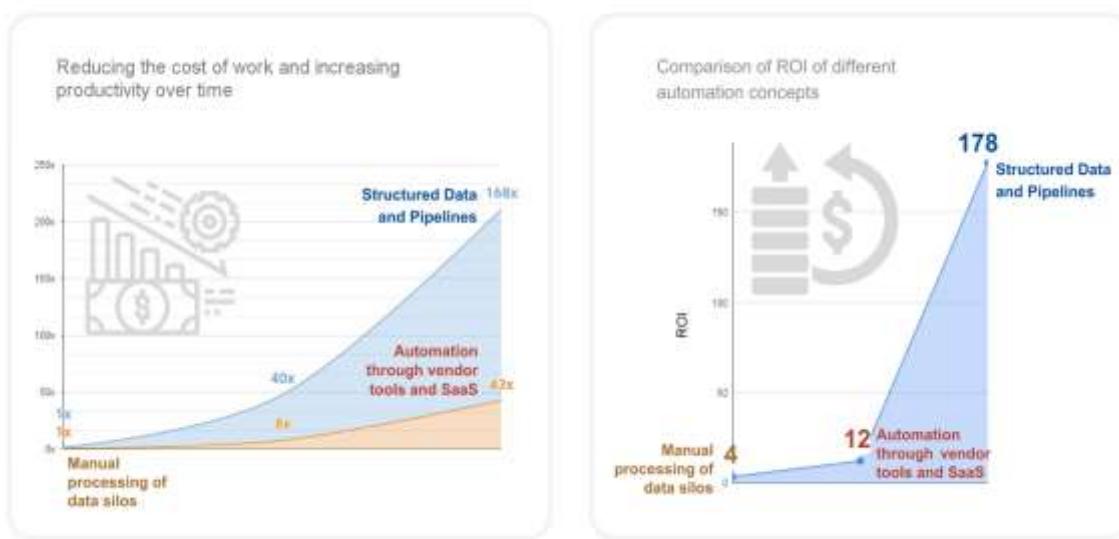


Fig. 7.3-2 ROI Pipeline datavalideringsprosessen reduserer kjøretiden med titalls og hundretalls ganger sammenlignet med behandling ved hjelp av klassiske verktøy [74].

I motsetning til manuelt arbeid i proprietære systemer (ERP, PMIS, CAD, osv.) kan du med pipelining øke hastigheten på oppgavene betydelig (fig. 7.3-2), unngå repeterende arbeid og automatisere oppstart av prosesser til rett tid (fig. 7.3-3).

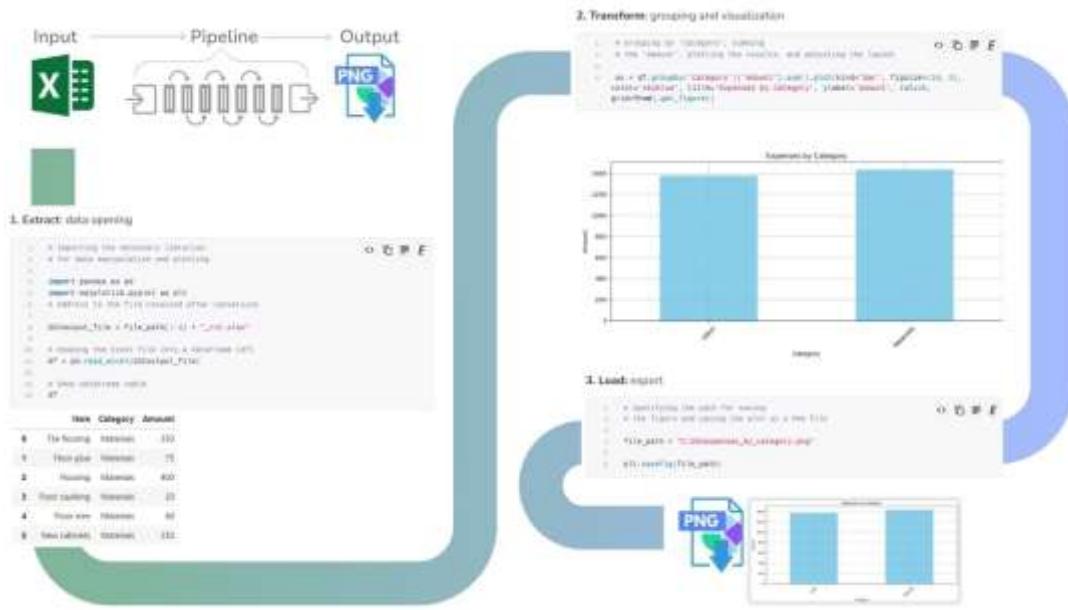


Fig. 7.3-3 ETL-eksempel Pipeline på automatisk innhenting av et diagram fra tabelldata i en XLSX-fil uten å åpne Excel.

For å behandle strømmedata og bygge en automatisert pipeline, på samme måte som i ETL-prosessen, må du identifisere datakildene på forhånd, samt tidsrammen for innsamling av data - enten for en spesifik forretningsprosess eller for hele selskapet.

I byggeprosjekter kommer data fra mange heterogene kilder med ulike oppdateringsintervaller. For å skape et pålitelig datautstilling vindu er det avgjørende å registrere når informasjonen hentes og oppdateres. Det gjør det mulig å ta beslutninger i tide og effektivisere prosjektstyringen.

Et alternativ er å starte sammenstillingsprosessen på et fast tidspunkt - for eksempel kl. 19.00, på slutten av arbeidsdagen. På dette tidspunktet aktiveres det første skriptet som er ansvarlig for å samle data fra ulike systemer og lagre (fig. 7.3-4 trinn 1). Deretter følger automatisert behandling og omforming av data til et strukturert format som egner seg for analyse (fig. 7.3-4 trinn 2-4). Til slutt genereres rapporter, dashboard og andre produkter som er beskrevet i de foregående kapitlene, automatisk ved hjelp av de forberedte dataene (fig. 7.3-4 trinn 6-7). Resultatet er at lederne allerede kl. 05.00 om morgenen har oppdaterte prosjektstatusrapporter i ønsket format (fig. 7.3-5).

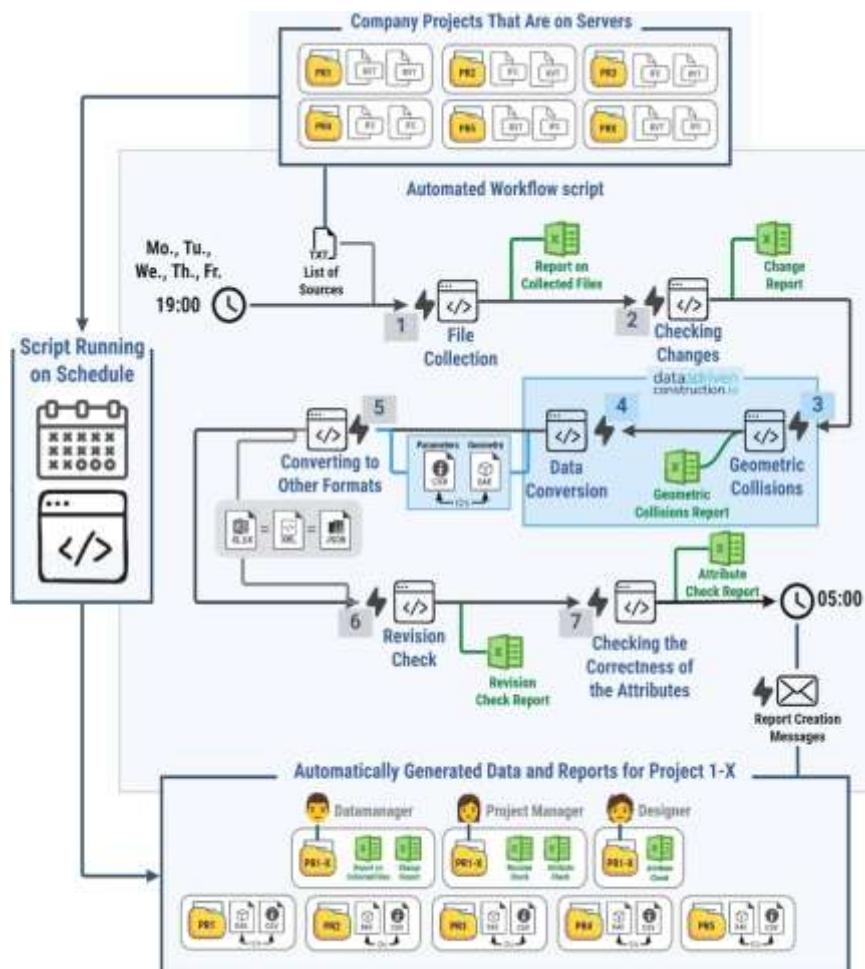


Fig. 7.3-4 Data i Pipeline, som samles inn automatisk om kvelden, behandles i løpet av natten slik at lederne har oppdaterte rapporter og ferske rapporter om morgenen.

Datainnsamling til rett tid, definisjon av KPI-er, automatisering av transformasjonsprosesser og visualisering gjennom dashboard er nøkkelementer i vellykket datadrevet beslutningstaking.

Slike automatiserte prosesser (fig. 7.3-4) kan kjøres helt autonomt: De kjører etter en tidsplan, behandler data uten tilsyn og kan distribueres enten i skyen eller på bedriftens egen server (fig. 7.3-5). Dette gjør det mulig å integrere slike ETL-rørledninger i den eksisterende IT-infrastrukturen, samtidig som man beholder kontrollen over dataene og får fleksibilitet når det gjelder skalering.

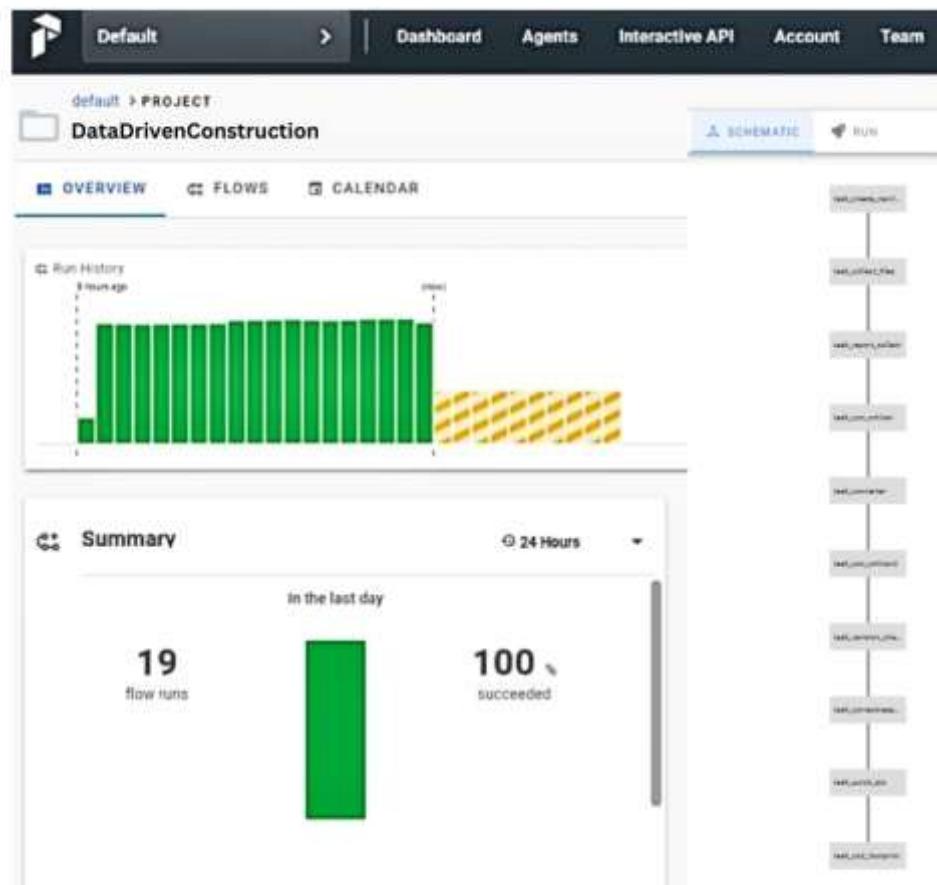


Fig. 7.3-5 Automatisk ETL -conveyor prosesser (fig. 7.3-4) på Prefect-plattformen, der 10 python-skript kjøres vekselvis etter kl. 19.00 hver arbeidsdag.

Automatisering av arbeidsflyter øker ikke bare teamets produktivitet ved å frigjøre tid til mer meningsfylte og mindre rutinepregede oppgaver, men fungerer også som et viktig første skritt i retning av å innlemme kunstig intelligens (AI) i forretningsprosessene, noe vi kommer nærmere inn på i kapittelet om prediktiv analyse og maskinlæring.

Pipeline -ETL datavalideringsprosess med LLM

I de foregående kapitlene om å lage datakrav og automatisere ETL, gikk vi steg for steg gjennom prosessen med klargjøring, transformasjon, validering og visualisering av data. Disse aktivitetene ble implementert som separate kodeblokker (figur 7.2-18 - figur 7.2-20), som hver utfører en spesifikk oppgave.

Nå har vi neste mål - å kombinere disse elementene i én enkelt, sammenhengende og automatisert pipeline for databehandling - en pipeline, ETL-Pipeline - der alle trinnene (innlasting, validering, visualisering, eksport) utføres sekvensielt i ett enkelt skript som kan kjøres automatisk.

I det følgende eksemplet vil en komplett databehandlingssyklus bli implementert: fra innlasting av CSV-kildefilen → til kontroll av struktur og verdier ved hjelp av regulære uttrykk → beregning av resultatene → generering av en visuell rapport i PDF-format.

- ☞ Du kan bruke følgende tekstforespørsel til LLM. for å hente den aktuelle koden:

Vennligst skriv et kodeeksempel som laster inn data fra CSV, validerer DataFrame ved hjelp av regulære uttrykk, sjekker identifikatorer i formatet 'W-NEW' eller 'W-OLD', energieffektivitet med bokstavene 'A' til 'G', garantiperiode og utskiftingssyklus med numeriske verdier i år og til slutt oppretter en rapport med en opptelling av godkjente og mislykkede verdier, genererer en PDF med et histogram av resultater og legger til en tekstbeskrivelse. ↴

💡 LLMs svar :



```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$', 
14     'Energy_Efficiency': r'^[A-G]\+*$',
15     'Warranty_Period': r'^\d+\syears$',
16     'Replacement_Cycle': r'^\d+\syears$'
17 }
18
19 for column, pattern in patterns.items():
20     df[f'verified_{column}'] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38 # Add text page
39 plt.figure()
40 plt.axis('off')
41 text_str = "Verification Results Summary"
42 plt.text(0.1, 0.5, text_str, wrap=True)
43 pdf.savefig()
44 plt.close()
45
46 pdf_path
47

```

Fig. 7.3-6 Pipeline (ETL) automatiserer hele syklusen av databehandling: fra nedlasting og validering til opprettelse av en strukturert rapport i PDF-format .

Den automatiserte koden (fig. 7.3-6) i LLM-chatterommet eller i DIE vil, etter å ha kopiert koden, validere dataene fra CSV -filen ved hjelp av de angitte regulære uttrykkene, opprette en rapport om antall godkjente og underkjente poster, og deretter lagre valideringsresultatene som en PDF -fil.

Denne ETL -transportørstrukturen, der hvert trinn - fra datalasting til rapportgenerering - er implementert som en separat modul, sikrer gjennomsiktighet, skalerbarhet og reproducerbarhet. Ved å presentere valideringslogikken som en lettles Python-kode blir prosessen gjennomsiktig og forståelig, ikke bare for utviklere, men også for spesialister innen datahåndtering, kvalitet og analyse.

Pipelines -tilnærming til automatisering av databehandling gjør det mulig å standardisere prosesser, øke repeterbarheten og forenkle tilpasningen til nye prosjekter. Dette skaper en enhetlig metodikk for analyse av data, uansett kilde eller type oppgave - enten det dreier seg om samsvarstesting, generering av rapporter eller overføring av data til eksterne systemer.

Slik automatisering reduserer menneskelige feil, reduserer avhengigheten av proprietære løsninger og øker nøyaktigheten og påliteligheten til resultatene, noe som gjør dem egnet for både operativ analyse på prosjektnivå og strategisk analyse på bedriftsnivå.

Rørledning -ETL: verifisering av data og informasjon om prosjektelementer i CAD (BIM)

Data fra CAD-systemer og databaser (BIM) er noen av de mest sofistikerte og dynamisk oppdaterte datakildene i byggebransjen. Disse applikasjonene beskriver ikke bare prosjektet ved hjelp av geometri, men supplerer det også med flere lag med tekstlig informasjon: volumer, materialegenskaper, romtildelinger, energieffektivitetsnivåer, toleranser, forventet levetid og andre attributter.

Attributter som tilordnes enheter i CAD -modeller dannes på designstadiet og blir grunnlaget for videre forretningsprosesser, inkludert kostnadsberegning, planlegging, livssyklusvurdering og integrering med ERP og CAFM -systemer, der effektiviteten i prosessene i stor grad avhenger av kvaliteten på dataene som kommer fra designavdelingene.

Den tradisjonelle tilnærmingen til attributtvalidering i CAD- (BIM-) -modeller innebærer manuell validering (fig. 7.2-1), noe som blir en lang og kostbar prosess når modellvolumet er stort. Med tanke på volumet og antallet moderne byggeprosjekter og de regelmessige oppdateringene av disse, blir prosessen med datavalidering og -transformasjon uholdbar og uoverkommelig.

Hovedentrepreneur og prosjektledere står overfor behovet for å behandle store mengder prosjektdata, inkludert flere versjoner og fragmenter av de samme modellene. Dataene kommer fra prosjekteringsorganisasjoner i RVT-, DWG-, DGN-, IFC-, NWD- og andre formater (figur 3.1-14) må jevnlig kontrolleres for å sikre at de er i samsvar med bransje- og selskapsstandarer.

Avhengigheten av manuelle handlinger og spesialisert programvare gjør datavalideringsprosessen til en flaskehals i arbeidsflyter knyttet til data fra bedriftsomfattende modeller. Automatisering og bruk av

strukturerte krav kan eliminere denne avhengigheten, noe som øker hastigheten og påliteligheten i datavalideringen dramatisk (figur 7.3-7).

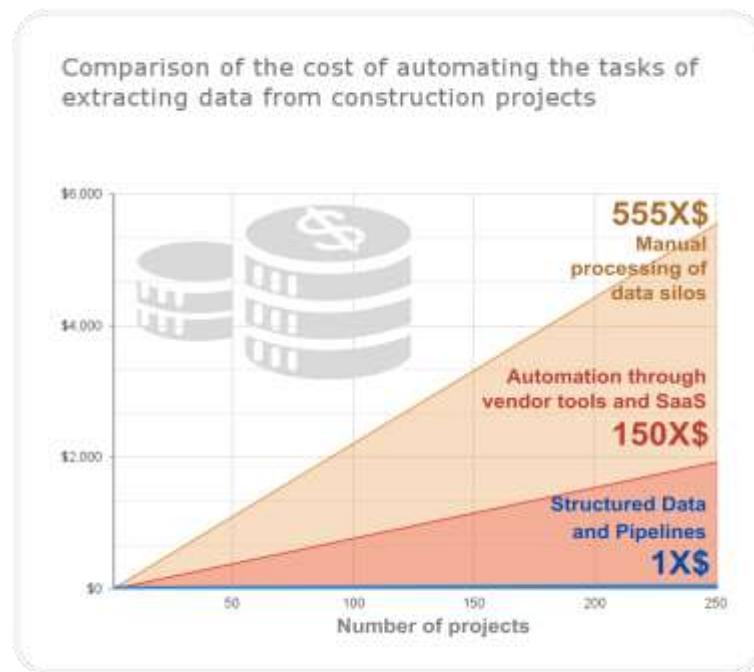
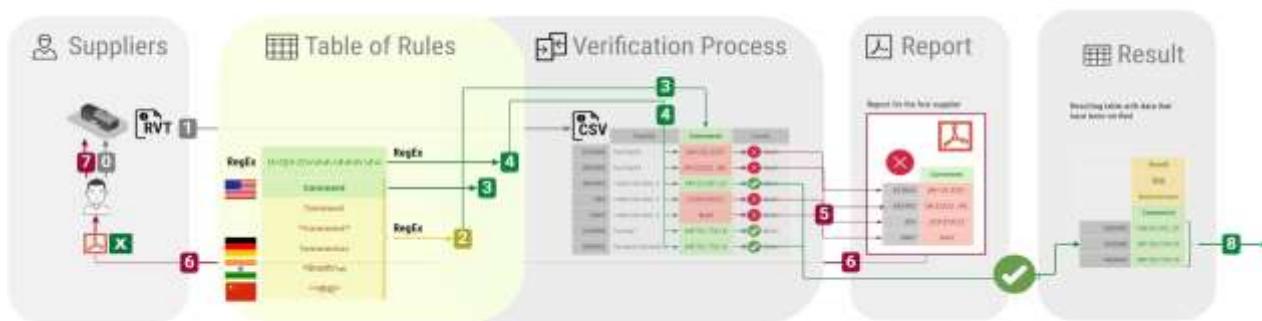


Fig. 7.3-7 Automatisering øker hastigheten på dataverifikasiing og -behandling, noe som reduserer arbeidskostnadene med flere titalls ganger [140].

CAD-datavalideringsprosessen omfatter datauttrekk (ETL-stadiet Extract) fra ulike lukkede (RVT, DWG, DGN, NWS, etc.) eller åpne semistrukturerte og parametriske formater (IFC, CPXML, USD...) eller åpne semistrukturerte og parametriske formater (IFC, CPXML, USD), der regeltabeller kan brukes på hvert attributt og dets verdier (Transform-trinn) ved hjelp av regulære uttrykk RegEx (fig. 7.3-8), en prosess som vi diskuterte i detalj i bokens fjerde del.

Opprettelsen av en PDF-feilrapport og validerte poster bør fullføres med utdata (Load step) i strukturerte formater som kun tar hensyn til validerte enheter som kan brukes i videre prosesser.



Figur 7.3-8 Datavalideringsprosess fra prosjektdataleverandører til sluttrapport validert ved hjelp av regulære uttrykk.

Automatisering av validering av data fra CAD-systemer (BIM) med strukturerte krav og strømming av nye data som behandles gjennom ETL-Pipelines (figur 7.3-9), reduserer behovet for manuell involvering i valideringsprosessen (hver av validerings- og datakravprosessene har blitt diskutert i tidligere kapitler).

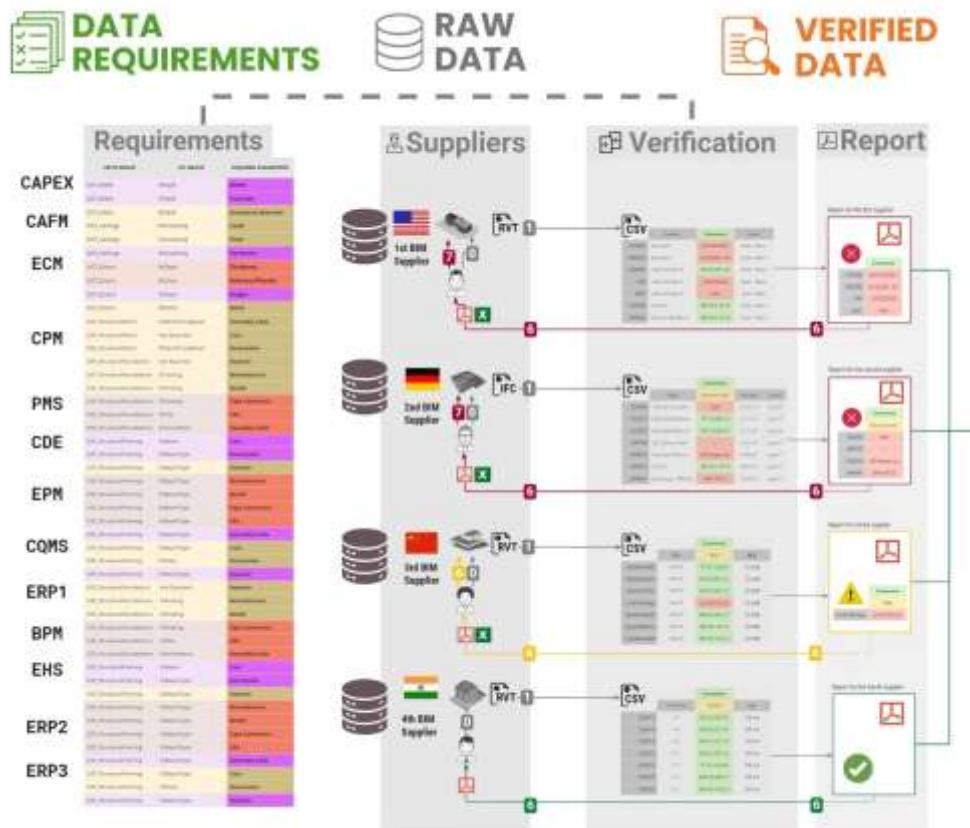


Fig. 7.3-9 Automatisering av datavalidering ved hjelp av ETL forenkler ledelsen av byggeprosjekter ved at prosessene går raskere.

Tradisjonelt kan validering av modeller levert av entreprenører og CAD (BIM)-spesialister ta dager eller uker. Med innføringen av automatiserte ETL-prosesser kan dette imidlertid reduseres til noen få minutter. I en typisk situasjon sier entreprenøren *"Modellen er validert og i samsvar med kravene."* Denne uttalelsen starter en kjede av verifisering av entreprenørens påstand om datakvalitet:

- ⌚ Prosjektleder - *"Entreprenøren sier: 'Modellen er testet, alt er i orden'."*
- ⌚ Data Manager - Last inn validering:
 - ⌚ Et enkelt skript i Pandas oppdager brudd på få sekunder. Automatisering eliminerer tvister:
 - ⌚ Category: OST_StructuralColumns, Parameter: FireRating IS NULL.
 - ⌚ Generer en liste over brudd-ID-er→ eksporter til Excel/PDF.

Et enkelt skript i Pandas oppdager innbruddet på få sekunder:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Filtrering  
issues = df[df["FireRating"].isnull()] # Tomme verdier  
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Eksporter ID-er
```

- ☛ Data Manager til Project Manager - "*En sjekk av viser at 18 kolonner ikke har FireRating-parameteren fylt ut.*"
- ☛ Prosjektleder til entreprenør - "*Modellen returneres for revisjon: FireRating-parameteren er obligatorisk, uten den er det umulig å godta den*"

Resultatet er at DAK-modellen ikke gjennomgår validering, automatisering eliminerer tvister, og entreprenøren mottar nesten umiddelbart en strukturert rapport med en liste over ID-er for problematiske elementer. På denne måten blir valideringsprosessen transparent, repeterbar og beskyttet mot menneskelige feil (fig. 7.3-10).

Denne tilnærmingen gjør datavalideringsprosessen til en ingeniørfunksjon i stedet for manuell kvalitetskontroll. Dette øker ikke bare produktiviteten, men gjør det også mulig å bruke den samme logikken på alle selskapets prosjekter, noe som muliggjør en ende-til-ende digital transformasjon av prosesser, fra design til drift.

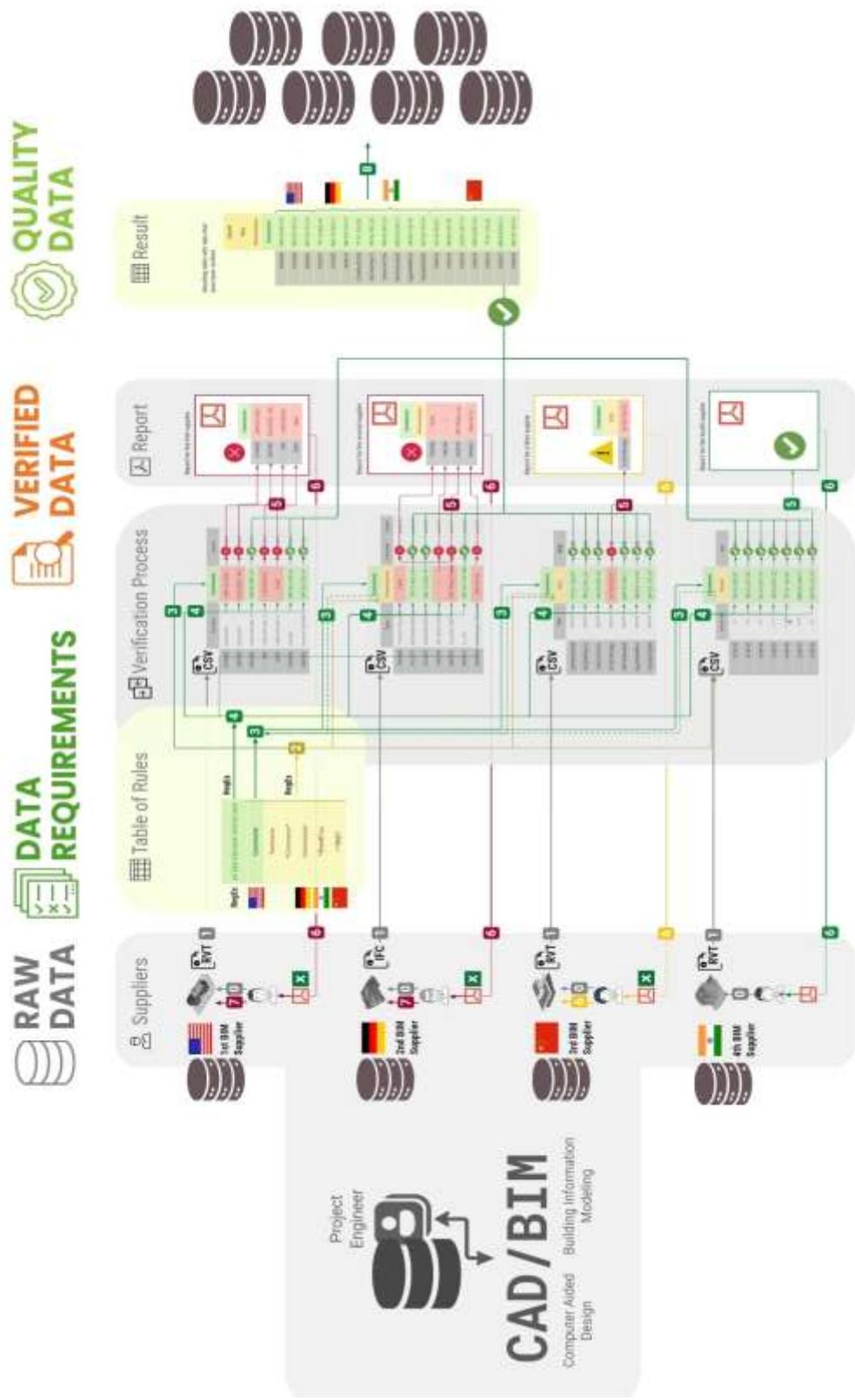


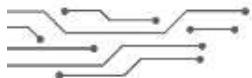
Fig. 7.3-10 Automatisering av kontrollen av elementattributter eliminerer menneskelige feil og reduserer sannsynligheten for feil.

Ved hjelp av automatiserte pipelines (fig. 7.3-10) kan systembrukere som forventer kvalitetsdata fra CAD-(BIM-) -systemer, umiddelbart få de utdataene de trenger - tabeller, dokumenter, bilder - og raskt integrere dem i arbeidsoppgavene sine.

Automatiseringen av kontroll, prosessering og analyse fører til en endring i måten man tilnærmer seg styring av byggeprosjekter på, spesielt interoperabilitet mellom ulike systemer, uten bruk av komplekse og dyre modulære proprietære systemer eller lukkede leverandørløsninger.

Mens konsepter og markedsføringsakronymer kommer og går, vil selve valideringsprosessene for datakrav for alltid forblи en integrert del av forretningsprosessene. I stedet for å lage flere og flere spesialiserte formater og standarder, bør byggebransjen se til verktøy som allerede har vist seg å være effektive i andre bransjer. I dag finnes det kraftige plattformer for automatisering av databehandling og prosessintegrasjon som gjør det mulig for bedrifter å redusere tiden som brukes på rutineoperasjoner, og minimere feil i Extract, Transform og Load.

Et av de mest populære eksemplene på løsninger for automatisering og orkestrering av ETL -prosesser er Apache Airflow, som gjør det mulig å organisere komplekse databehandlingsprosesser og administrere ETL-pipelines. I tillegg til Airflow brukes også andre lignende løsninger som Apache NiFi for dataruting og -strømming og n8n for automatisering av forretningsprosesser.



KAPITTEL 7.4.

ORKESTRERING AV ETL OG ARBEIDSFLYTER: PRAKTISKE LØSNINGER

DAG og Apache Airflow: automatisering og orkestrering av arbeidsflyt

Apache Airflow er en gratis plattform med åpen kildekode, som er utviklet for å automatisere, orkestrere og overvåke arbeidsflyter (ETL -conveyors).

Hver dag er det nødvendig å jobbe med store datamengder:

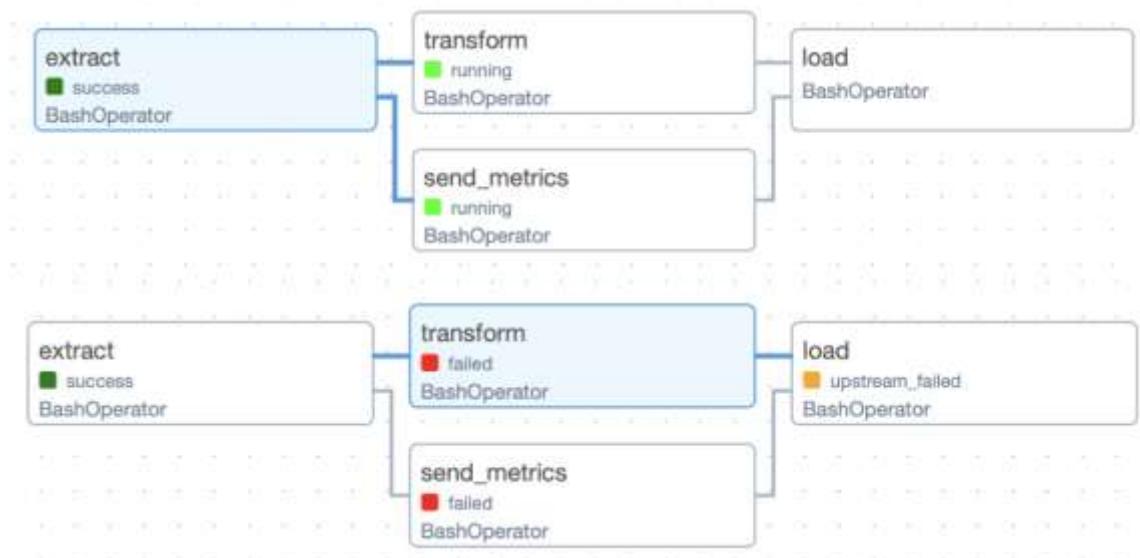
- Last ned filer fra ulike kilder - Trekk utfor eksempel fra leverandører eller kunder).
- Transformere disse dataene til ønsket format - Transform (strukturere, rense og validere)
- Send resultater for verifisering og opprett rapporter - Last inn (last opp til nødvendige systemer, dokumenter, databaser eller dashbord).

Manuell kjøring av slike ETL-prosesser tar mye tid og medfører risiko for menneskelige feil. En endring i datakilden eller en feil i ett av trinnene kan føre til forsinkelser og feilaktige resultater.

Automatiseringsverktøy, som Apache Airflow, gjør det mulig å bygge en pålitelig ETL -transportør, minimere feil, redusere behandlingstiden og sikre at dataene er korrekte i alle ledd. Kjernen i Apache Airflow er konseptet DAG (Directed Acyclic Graph) - en rettet acyklistisk graf der hver oppgave (operatør) er koblet til andre avhengigheter og utføres strengt i en spesifisert rekkefølge. DAG eliminerer sykluser, noe som gir en logisk og forutsigbar struktur for oppgaveutførelse.

Airflow tar seg av orkestreringen - håndtering av avhengigheter mellom oppgaver, kontroll av utførelsesplaner, sporing av status og automatisk reaksjon på feil. Denne tilnærmingen minimerer manuelle inngrep og sikrer påliteligheten til hele prosessen.

Task Orchestrator er et verktøy eller system som er utviklet for å administrere og kontrollere oppgaveutførelse i komplekse databehandlings- og informasjonsmiljøer. Det forenkler prosessen med å distribuere, automatisere og administrere oppgaveutførelse for å forbedre ytelsen og optimalisere ressursene.



Figur 7.4-1 Apache Airflow har et brukervennlig grensesnitt der du kan visualisere DAG -ETL, se kjøringslogger, status for oppstart av oppgaver med mer.

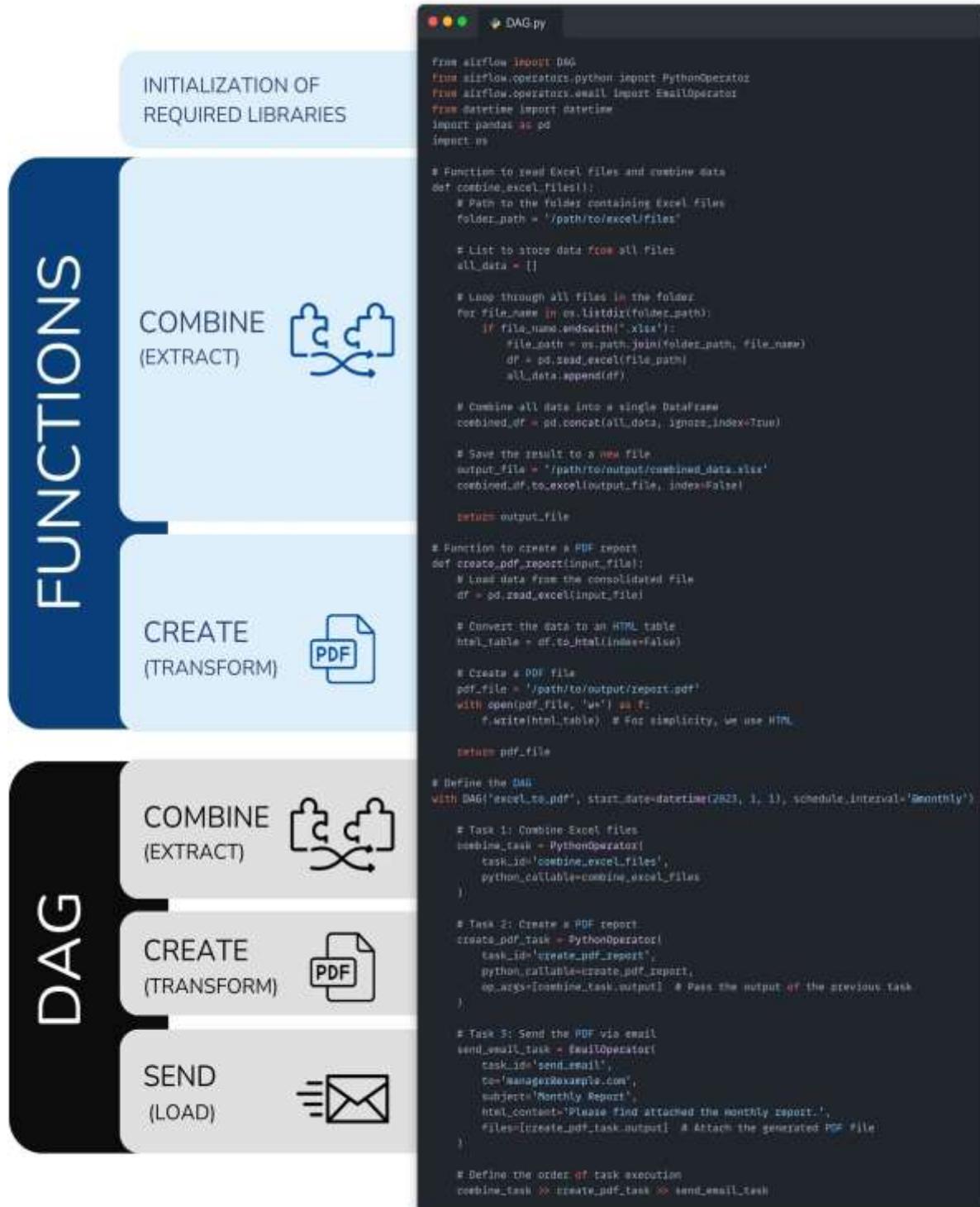
Airflow er mye brukt til orkestrering og automatisering av distribuert databehandling, databehandling, ETL (Extract, Transform, Load), prosessstyring, oppgaveplanlegging og andre datascenarier. Som standard bruker Apache Airflow SQLite som database.

Et eksempel på en enkel DAG, som ligner på ETL, består av oppgaver - Extract, Transform og Load. I grafen, som styres via brukergrensesnittet (fig. 7.4-1), defineres rekkefølgen på utførelsen av oppgavene (kodefragmenter): For eksempel utføres extract først, deretter transform (og sending_metrics), og load-oppgaven fullfører arbeidet. Når alle oppgavene er fullført, anses datalastingsprosessen som vellykket.

Apache Airflow: praktisk anvendelse ETL-automatisering

Apache Airflow er mye brukt til å organisere komplekse databehandlingsprosesser, noe som gjør det mulig å bygge fleksible ETL -transportører. Apache Airflow kan kjøres enten via et webgrensesnitt eller programmatisk ved hjelp av Python-kode (fig. 7.4-2). I webgrensesnittet (fig. 7.4-3) kan administratorer og utviklere visuelt overvåke DAG-er, kjøre oppgaver og analysere kjøringsresultater.

Ved hjelp av DAG kan du angi en klar rekkefølge for oppgavene, håndtere avhengigheter mellom dem og automatisk reagere på endringer i kildedataene. La oss se på et eksempel på bruk av Airflow for å automatisere rapporteringsbehandling (fig. 7.4-2).



Figur 7.4-2 ETL-konsept -transportør for databehandling ved hjelp av Apache Airflow.

Dette eksempelet (figur 7.4-2) tar for seg DAG, som utfører sentrale oppgaver i ETL -transportøren:

ETL (EXTRACT, TRANSFORM, LOAD) PIPELINE

■ **Les Excel -filer (Extract):**

- Sekvensiell gjennomgang av alle filer i en gitt katalog.
- Les data fra hver fil ved hjelp av pandas-biblioteket.
- Slå sammen alle data til en enkelt DataFrame.

■ **Opprett PDF -dokument (Transform):**

- Transformer den sammenslåtte DataFrame til en HTML -tabell.
- Lagre tabellen som PDF (i demoversjonen - via HTML).

■ **Sende en rapport via e-post (Last inn):**

- Bruk EmailOperator for å sende PDF -dokument via e-post.

■ **Konfigurer DAG:**

- Definere rekkefølgen på oppgavene: hente ut data → generere rapport → sende.
- Tilordne en lanseringsplan (@monthly - første dag i hver måned).

Det automatiserte ETL -eksemplet (figur 7.4-2) viser hvordan man samler inn data fra Excel -filer, lager et PDF -dokument og sender det via e-post. Dette er bare ett av mange mulige bruksområder for Airflow. Dette eksemplet kan tilpasses enhver spesifikk oppgave for å forenkle og automatisere databehandlingen.

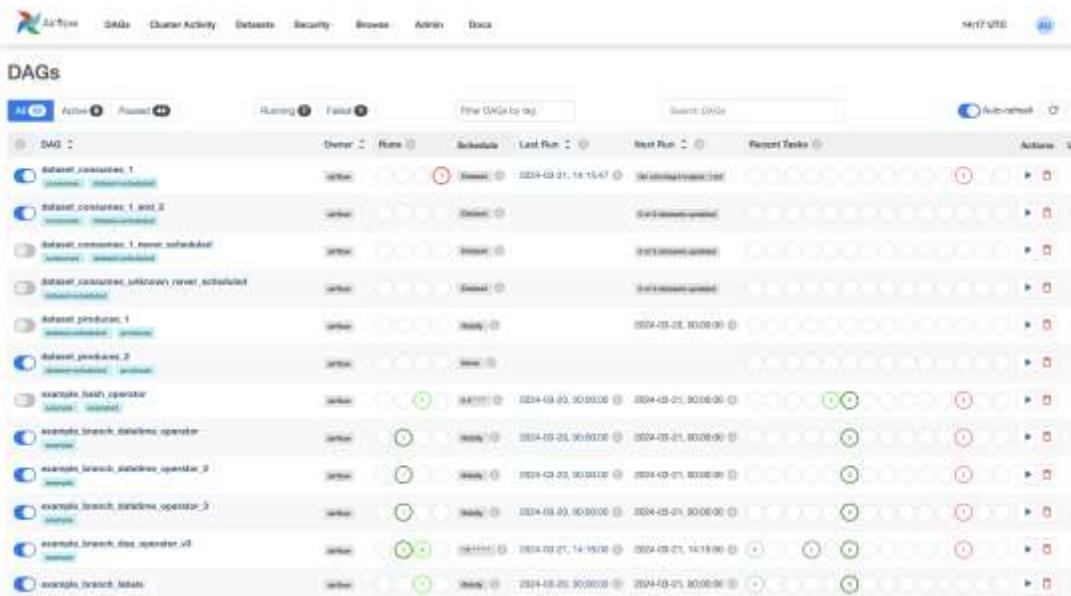


Fig. 7.4-3 Oversikt over alle DAG-er i miljøet med informasjon om de siste kjøringene.

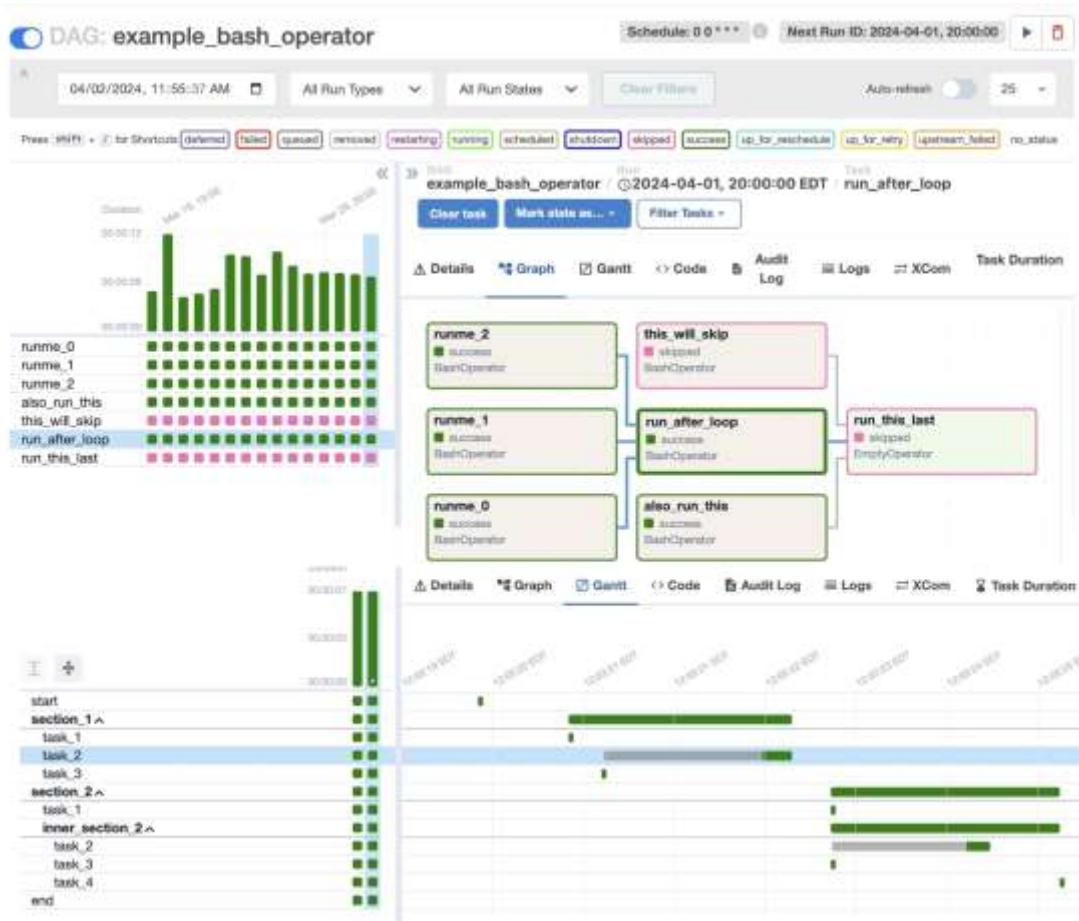
Apache Airflow-nettgrensesnittet (figur 7.4-3) gir et omfattende visuelt miljø for administrasjon av dataarbeidsflyter. Det viser DAG-er som interaktive grafer der noder representerer oppgaver og kanter representerer avhengigheter mellom dem, noe som gjør det enkelt å spore komplekse dataarbeidsflyter. Grensesnittet inkluderer et dashbord med informasjon om status for oppgaveutførelse, kjøringshistorikk, detaljerte logger og ytelsesmålinger. Administratører kan starte oppgaver manuelt, starte mislykkede operasjoner på nytt, suspendere DAG-er og konfigurere miljøvariabler, alt gjennom et intuitivt brukergrensesnitt.

En slik arkitektur kan suppleres med datavalidering, varslinger om kjøringsstatus, integrasjon med eksterne

API-er eller databaser. Med Airflow kan du tilpasse DAG på en fleksibel måte: legge til nye oppgaver, endre rekkefølgen på dem, kombinere kjeder - noe som gjør det til et effektivt verktøy for automatisering av komplekse databehandlingsprosesser. Når du kjører DAG i Airflow-nettgrensesnittet (fig. 7.4-3, fig. 7.4-4), kan du overvåke statusen til oppgavene. Systemet bruker fargeindikasjon:

- Grønn - oppgaven er fullført.
- Gul - prosessen er i gang.
- Rød - en feil under utførelsen av oppgaven.

I tilfelle feil (f.eks. manglende fil eller ødelagt datastruktur), sender systemet automatisk et varsel.



Figur 7.4-4 Apache Airflow forenkler problemdiagnosering, prosessoptimalisering og teamsamarbeid i forbindelse med komplekse databehandlingsrørledninger.

Apache Airflow er praktisk fordi det automatiserer rutineoppgaver, slik at du slipper å utføre dem manuelt. Det gir pålitelighet ved å overvåke prosessutførelsen og varsle om feil umiddelbart. Systemets fleksibilitet gjør det enkelt å legge til nye oppgaver eller endre eksisterende oppgaver, slik at arbeidsflyten kan tilpasses endrede krav.

I tillegg til Apache Airflow finnes det lignende verktøy for orkestrering av arbeidsflyter. For eksempel tilbyr

Prefect (figur 7.3-5), som er åpen kildekode og gratis, en enklere syntaks og integreres bedre med Python, Luigi, som er utviklet av Spotify, tilbyr lignende funksjonalitet og fungerer godt med stordata. Det er også verdt å merke seg Kronos og Dagster, som tilbyr moderne tilnærmingar til å bygge Pipeline med fokus på modularitet og skalerbarhet. Valget av oppgaveorkestreringsverktøy avhenger av prosjektets spesifikke behov, men de bidrar alle til å automatisere komplekse ETL dataprosesser.

Apache NiFi, en plattform med åpen kildekode, som er utviklet for strømming og ruting av data. I motsetning til Airflow, som fokuserer på batchbehandling og avhengighetsstyring, fokuserer NiFi på sanntids datatransformasjon og fleksibel ruting mellom systemer.

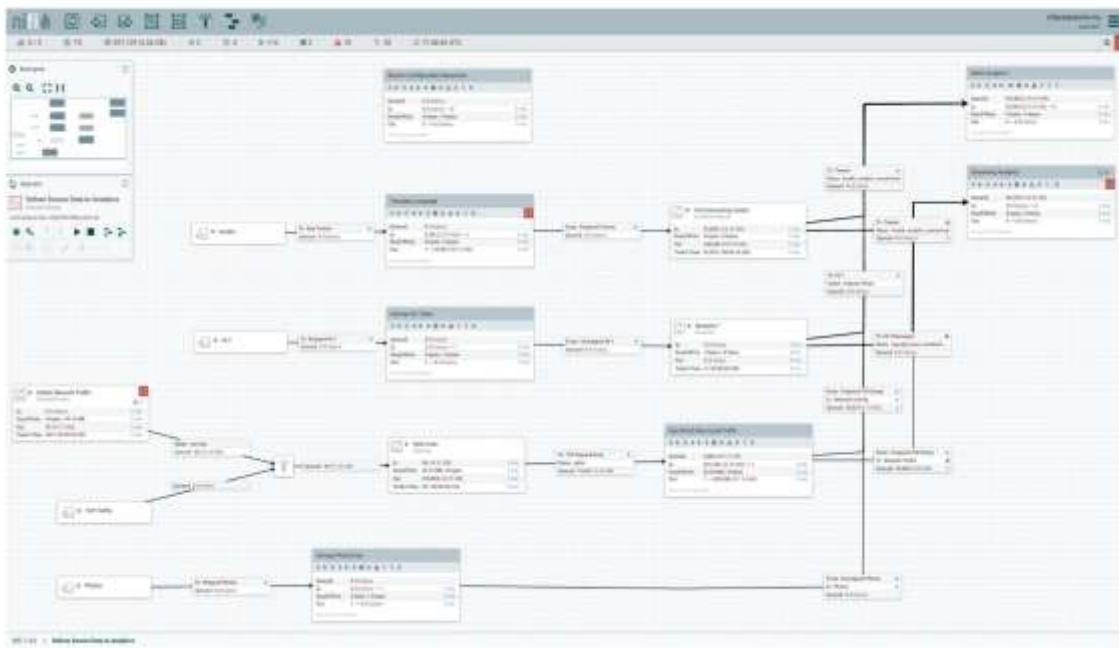
Apache NiFi for ruting og datakonvertering

Apache NiFi er en kraftig plattform med åpen kildekode, som er utviklet for å automatisere dataflyten mellom ulike systemer. Den ble opprinnelig utviklet i 2006 av det amerikanske National Security Agency (NSA) under navnet "Niagara Files" for internt bruk. I 2014 ble prosjektet gitt åpen kildekode og overført til Apache Software Foundation, og ble en del av deres teknologioverføringsinitiativ [141].

Apache NiFi er utviklet for å samle inn, behandle og overføre data i sanntid. I motsetning til Airflow, som arbeider med batchoppgaver og krever veldefinerte tidsplaner, opererer NiFi i en strømbehandlingsmodus som gjør det mulig å overføre data kontinuerlig mellom ulike tjenester.

Apache NiFi er ideell for integrasjon med IoT -enheter, byggensensorer, overvåkingssystemer og f.eks. strømmevalidering av CAD -formater på en server der det kan være behov for umiddelbar respons på endringer i data.

Med innebygde filtrerings-, transformerings- og ratingsverktøy gjør NiFi det mulig å standardisere data (Transform) før de overføres (Load) til lagrings- eller analysesystemer. En av de viktigste fordelene er den innebygde sikkerhetsstøtten og tilgangskontrollen, noe som gjør den til en pålitelig løsning for håndtering av sensitiv informasjon.



Figur 7.4-5 Grafisk fremstilling av dataflyten i Apache NiFi-grensesnittet.

Apache NiFi håndterer effektivt strømming, filtrering og ruting av data i sanntid. Den er ideell for teknisk intensive scenarier der stabil informasjonsoverføring mellom systemer og høy gjennomstrømning er viktig.

Men når hovedmålet er å integrere ulike tjenester, automatisere rutineoperasjoner og raskt sette opp arbeidsflyter uten dyptgående programmeringskunnskaper, er det behov for løsninger med lav inngangstterskel og maksimal fleksibilitet. Et slikt verktøy er n8n - en Low-Code /No-Code -klassplattform med fokus på forretningsautomatisering og visuell orkestrering av prosesser.

n8n Low-Code, No-Code prosessorkestrering

n8n er en Open Source Low-Code / No-Code plattform for å bygge automatiserte arbeidsflyter, som kjennetegnes av brukervennlighet, fleksibilitet og muligheten til raskt å integrere med et bredt spekter av eksterne tjenester.

No-Code er en metode for å skape digitale produkter uten å skrive kode. Alle elementer i prosessen - fra logikk til grensesnitt - realiseres utelukkende ved hjelp av visuelle verktøy. No-Code-plattformer er rettet mot brukere uten teknisk bakgrunn og gjør det mulig å raskt lage automatiseringer, skjemaer, integrasjoner og webapplikasjoner. Eksempel: En bruker setter opp automatisk sending av varsler eller Google Sheets-integrasjon ved hjelp av et dra-og-slipp-grensesnitt uten programmeringskunnskaper.

Med åpen kildekode og lokale distribusjonsmuligheter gir n8n i automatiserings- og ETL Pipeline-opprettelsesprosesser selskaper full kontroll over dataene sine, samtidig som det sørger for sikkerhet og uavhengighet fra skyleverandører.

I motsetning til Apache Airflow, som er rettet mot beregningsoppgaver med rigid orkestrering og krever kunnskaper i Python, tilbyr n8n en visuell editor som gjør det mulig å lage skript uten å ha kjennskap til programmeringsspråk (figur 7.4-6). Grensesnittet gjør det mulig å opprette automatiserte prosesser uten å skrive kode (No-Code), men i mer komplekse scenarier kan brukerne legge til sine egne JavaScript og Python -funksjoner for å utvide funksjonene (Low-Code).

Low-Code er en tilnærming til programvareutvikling der den grunnleggende logikken i en applikasjon eller prosess skapes ved hjelp av et grafisk grensesnitt og visuelle elementer, og der programkoden kun brukes til å tilpasse eller utvide funksjonaliteten. Low-Code-plattformer gjør det mulig å fremskynde utviklingen av løsninger betydelig ved å involvere ikke bare programmerere, men også forretningsbrukere med grunnleggende tekniske ferdigheter. Eksempel: En bruker kan bygge en forretningsprosess fra ferdige blokker og, om nødvendig, legge til sitt eget skript i JavaScript eller Python.

Selv om n8n er posisjonert som en plattform med lav inngangstterskel, kreves det grunnleggende programmeringskunnskaper, forståelse av webteknologi og ferdigheter i å jobbe med API. Systemets fleksibilitet gjør at det kan tilpasses et bredt spekter av oppgaver - fra automatisert databehandling til integrasjon med messengers, IoT -enheter og skytjenester.

Viktige funksjoner og fordeler ved bruk av n8n:

- **Åpen kildekode** og lokale distribusjonsalternativer sikrer fullstendig datakontroll, overholdelse av sikkerhetskrav og uavhengighet av skyleverandører.
- **Integrasjon med over 330 tjenester**, inkludert CRM, ERP, e-handel, skyplattformer, messengers og databaser.
- **Fleksible scenarier**: fra enkle varsler til komplekse kjeder med API-behandling av -forespørsler, beslutningslogikk og tilkobling av AI -tjenester.
- **Støtte for JavaScript og Python**: Brukerne kan legge inn tilpasset kode etter behov, noe som utvider automatiseringsmulighetene.
- **Intuitivt visuelt grensesnitt**: gjør at du raskt kan konfigurere og visualisere alle prosesstrinn.

Low-Code-klassenplattformer tilbyr verktøy for å skape digitale løsninger med minimalt med kode, noe som gjør dem ideelle for team som ikke har dyp teknisk ekspertise, men som trenger å automatisere prosesser.

I bygg- og anleggsbransjen kan n8n brukes til å automatisere en rekke prosesser, for eksempel integrering med prosjektstyringssystemer, strømningskontroll, skriving av ferdige rapporter og brev, automatisk oppdatering av materialbeholdningsdata, sending av statusmeldinger til teamene og mye mer. En spesialtilpasset Pipeline i n8n kan redusere manuelle operasjoner med flere ganger, redusere sannsynligheten for feil og fremskynde beslutningstaking for prosjektgjennomføring.

Du kan velge mellom nesten to tusen ferdige, gratis n8n Pipeline med åpen kildekode, tilgjengelig på: n8n.io/workflows, for å automatisere både arbeidsflyter i bygg og anlegg og personlige oppgaver, og redusere rutinemessige operasjoner.

Ta en av de ferdige Pipeline-malene, som er gratis tilgjengelig på n8n.io [142], som automatisk lager utkast

til svar i Gmail (fig. 7.4-6), noe som hjelper brukere som mottar mange e-poster eller har problemer med å skrive svar.

Denne n8n-malen "Gmail AI Auto-Responder: Create draft responses to incoming emails" (figur 7.4-6) analyserer innkommende e-postmeldinger ved hjelp av LLM fra ChatGPT, avgjør om det er behov for et svar, genererer et utkast fra ChatGPT og konverterer teksten til HTML og legger den til i meldingskjeden i Gmail. E-posten sendes ikke automatisk, slik at du kan redigere og godkjenne svaret manuelt. Oppsettet tar omtrent 10 minutter og inkluderer OAuth-konfigurasjon av Gmail API og OpenAI API-integrasjon. Resultatet er en praktisk og gratis løsning for å automatisere rutinemessig e-postkommunikasjon uten å miste kontrollen over innholdet i e-postene.

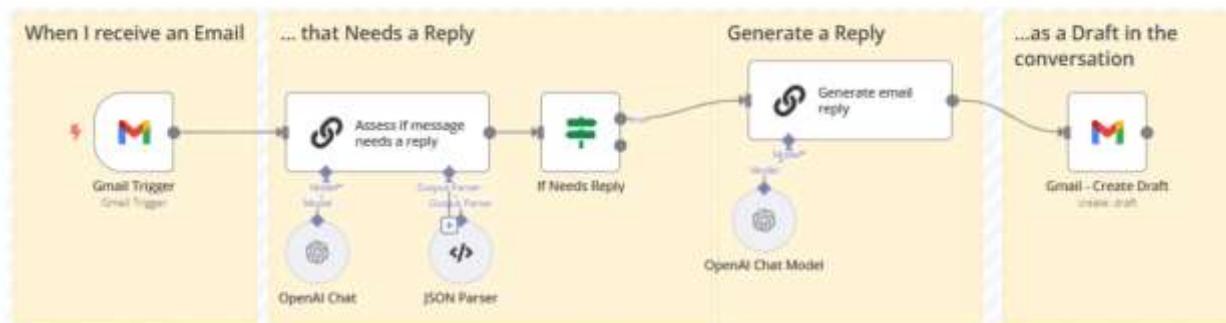


Fig. 7.4-6 Automatisert prosess for generering av e-postsvar ved hjelp av n8n.

Et annet eksempel på automatisering med n8n er å finne gode tilbud i eiendomsmarkedet [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", samler daglig inn relevante tilbud som samsvarer med de gitte kriteriene ved hjelp av Zillow API. Den beregner automatisk viktige investeringsmålinger (kontantavkastning, månedlig kontantstrøm, forskuddsbetaling), oppdaterer Google Sheets og sender en oppsummeringsrapport til e-post (figur 7.4-7), slik at investorer kan spare tid og svare raskt på de beste tilbudene.

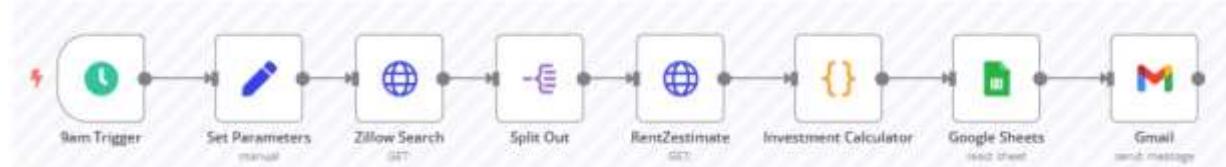


Fig. 7.4-7 Automatisert prosess for å vurdere hvor attraktiv eiendom er som investeringsobjekt.

Fleksibiliteten og utvidelsesmulighetene gjør n8n til et verdifullt verktøy for bedrifter som ønsker å gjennomføre digitale endringer og bli mer konkurransedyktige på markedet med relativt enkle og gratis verktøy med åpen kildekode.

Verktøy som Apache NiFi, Airflow og n8n kan betraktes som tre lag med databehandling (figur 7.4-8). NiFi administrerer dataflyten og sørger for levering og omforming, Airflow orkestrerer oppgaveutførelsen ved å samle data i prosesseringsrørledninger, og n8n automatiserer integrasjonen med eksterne tjenester og administrerer forretningslogikkken.



	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Fig. 7.4-8 Apache Airflow, Apache NiFi og n8n kan ses på som tre komplementære lag i en moderne datahåndteringsarkitektur.

Til sammen utgjør disse gratis verktøyene med åpen kildekode et eksempel på et effektivt økosystem for data- og prosesshåndtering i byggebransjen, som gjør det mulig for bedrifter å utnytte informasjon til å ta beslutninger og automatisere prosesser.

Neste steg: fra manuelle operasjoner til analysebaserte løsninger

Dagens byggfirmaer opererer i et miljø preget av stor usikkerhet: skiftende materialpriser, forsinkede leveranser, mangel på arbeidskraft og stramme prosjektfrister. Ved hjelp av analytiske dashbord, ETL - transportører og BI-systemer kan bedriftene raskt identifisere problemområder, vurdere ressurseffektiviteten og forutse endringer før de fører til økonomiske tap.

For å oppsummere denne delen er det verdt å fremheve de viktigste praktiske trinnene som vil hjelpe deg med å ta i bruk de omtalte teknologiene i det daglige arbeidet:

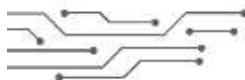
- Implementere datavisualiseringer og analytiske dashbord
 - Beherske prosessen med å lage dashbord for å overvåke viktige resultatindikatorer (KPI-er)
 - Bruk visualiseringsverktøy for dataene dine (Power BI, Tableau, Matplotlib, Plotly)
- Automatiser databehandlingen gjennom ETL -prosesser
 - Sett opp automatisk datainnsamling fra ulike kilder (dokumentasjon, tabeller, CAD) via ETL-prosesser
 - Organisere datatransformasjon (f.eks. kontroll av regulære uttrykk eller beregning) ved hjelp av Python-skript
 - Prøv å sette opp automatisk PDF- (eller DOC-) rapportering med FPDF-biblioteket, ved hjelp av data fra Excel-filer eller ved å trekke ut informasjon fra andre PDF-dokumenter
- Bruk språkmodeller (LLM) for automatisering
 - Bruke store språkmodeller (LLM) til å generere kode som hjelper deg med å hente ut og

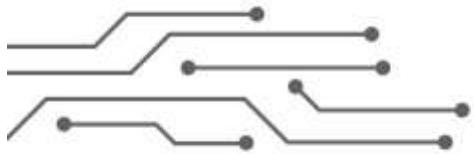
analysere data fra ustrukturerte dokumenter

- Gjør deg kjent med automatiseringsverktøyet n8n, og utforsk de ferdige malene og casestudiene på nettstedet deres. Identifiser hvilke prosesser i arbeidet ditt som kan automatiseres fullt ut ved hjelp av No-Code/Low-Code-tilnærmingen

En analytisk tilnærming til data- og prosessautomatisering reduserer ikke bare tiden som brukes på rutineoperasjoner, men forbedrer også kvaliteten på beslutningstakingen. Bedrifter som tar i bruk visuelle analyseverktøy og ETL -conveyors, får muligheten til å reagere raskt på endringer

Automatisering av forretningsprosesser ved hjelp av verktøy som n8n, Airflow og NiFi er bare det første skrittet mot digital modenhet. Det neste trinnet er kvalitetslagring og -håndtering av dataene som ligger til grunn for automatiseringen. I del 8 ser vi nærmere på hvordan byggefirmaer kan bygge opp en bærekraftig datalagringsarkitektur og gå fra et kaos av dokumenter og filer i flere formater til sentraliserte lagrings- og analyseplattformer.





VIII DEL

LAGRING OG HÅNDTERING AV DATA I BYGG- OG ANLEGGSBRANSJEN

Del 8 tar for seg moderne datalagrings- og datahåndteringsteknologier for byggebransjen. Den analyserer effektive formater for håndtering av store informasjonsmengder - fra enkle CSV og XLSX til de mer ytelsessterke Apache Parquet og ORC med en detaljert sammenligning av deres muligheter og begrensninger. Konseptene datavarehus (DWH), datasjøer) og hybridløsninger av disse (Data Lakehouse), samt prinsippene for datastyring) og dataminimalisme) diskuteres. Problemene med datasump) og strategier for å forhindre kaos i informasjonssystemer blir behandlet i detalj. Nye tilnæringer til dataarbeid presenteres, blant annet vektordatabaser og deres anvendelse i konstruksjon gjennom begrepet Bounding Box. Denne delen tar også for seg DataOps og VectorOps som nye standarder for organisering av dataarbeidsflyter.

KAPITTEL 8.1.

DATAINFRASTRUKTUR: FRA LAGRINGSFORMATER TIL DIGITALE REPOSITORIER

Dataatomer: grunnlaget for effektiv informasjonsforvaltning

Alt i universet består av de minste byggesteinene - atomer og molekyler, og over tid vender alle levende og ikke-levende ting uunngåelig tilbake til denne opprinnelige tilstanden. I naturen skjer denne prosessen med forbløffende hastighet, noe vi prøver å overføre til prosesser som styres av mennesker.

I skogen blir alle levende organismer etter hvert omdannet til en næringsrik substans som danner grunnlaget for nye planter. Disse plantene blir i sin tur mat for nye levende vesener som består av de samme atomene som skapte universet for millioner av år siden.

I forretningsverdenen er det også viktig å bryte ned komplekse, flerlagede strukturer til de mest grunnleggende, minimalt bearbeidede enhetene - omrent som atomer og molekyler i naturen. På denne måten kan atomer av data lagres og håndteres effektivt, slik at de blir til et rikt, fruktbart grunnlag som blir en nøkkelressurs for veksten i analysene og beslutningskvaliteten.



Fig. 8.1-1 Analyse og beslutningstaking er basert på gjenbrukte data som en gang har blitt behandlet og lagret.

Musikk består av toner som settes sammen til komplekse musikkstykker, mens ord skapes ut fra en primitiv enhet, bokstavlyden. Enten det dreier seg om natur, vitenskap, økonomi, kunst eller teknologi, viser verden en bemerkelsesverdig enhet og harmoni i sin jakt på ødeleggelse, struktur, sykling og nyskaping. På samme

måte brytes prosesser i kalkulasjonssystemer ned i små strukturerte enheter - ressurselementer - på kalkulasjonsnivå og tidsplaner. Disse enhetene, som sedler, brukes deretter til å danne mer komplekse kalkyler og tidsplaner. Det samme prinsippet brukes i systemer for datastøttet design, der komplekse arkitekt- og ingenørprosjekter bygges opp av grunnelementer - enkeltelementer og bibliotekskomponenter - som danner en komplett 3D-modell -modell av prosjektet til en kompleks bygning eller struktur.

Konseptet med syklisitet og struktur som er iboende i naturen og vitenskapen, gjenspeiles også i den moderne dataverdenen. På samme måte som alt levende i naturen går tilbake til atomer og molekyler, har informasjon en tendens til å gå tilbake til sin mest primitive form i moderne databehandlingsverktøy.

De minste elementene med sin endelige udelelighet er de grunnleggende byggsteinene i forretningsprosessene. Det er viktig å tenke nøye gjennom hvordan man samlar inn, strukturerer (bryter ned i atomer) og lagrer disse bittesmå byggsteinene fra ulike kilder. Organisering og lagring av data handler ikke bare om å bryte dem ned i sine bestanddeler. Det er like viktig å sørge for at de integreres og lagres på en strukturert måte, slik at dataene enkelt kan hentes frem, analyseres og brukes som beslutningsgrunnlag når det er behov for det.

For å kunne behandle informasjon effektivt må formatet og metodene for datalagring velges med omhu - på samme måte som jordsmonnet må forberedes for at trær skal vokse. Datalagrene må organiseres slik at informasjonen er av høy kvalitet og relevans, og overflødige eller irrelevante data må elimineres. Jo bedre denne "informasjonsjorden" er strukturert, desto raskere og mer nøyaktig kan brukerne finne de riktige dataene og løse analytiske problemer.

Lagring av informasjon: filer eller data

Datavarehus gjør det mulig for bedrifter å samle inn og kombinere informasjon fra ulike systemer og skape ett enkelt senter for videre analyse. De historiske dataene som samles inn, muliggjør ikke bare dypere analyser av prosesser, men også identifisering av mønstre som kan påvirke virksomhetens resultater.

La oss si at et selskap jobber med flere prosjekter samtidig. En ingeniør ønsker å finne ut hvor mye betong som er støpt og hvor mye som gjenstår å kjøpe. Med en tradisjonell tilnærming ville han måtte søke manuelt på serveren og åpne flere estimattabeller, sammenligne dem med sertifikatene for fullført arbeid og sjekke den aktuelle lagersaldoen. Dette tar timer eller til og med dager. Selv med ETL-prosesser og automatiske skript forblir oppgaven halvmanuell: Ingeniøren må fortsatt spesifisere banen til mapper eller spesifikke filer på serveren manuelt. Dette reduserer den samlede effekten av automatiseringen, ettersom det fortsatt tar verdifull arbeidstid.

Når man går over til datahåndtering, får ingeniøren tilgang til en enhetlig lagringsstruktur der informasjonen oppdateres i sanntid i stedet for å jobbe med serverens filsystem. En enkelt spørring - i form av kode, SQL - forespørsler eller til og med et LLM -agentanrop - kan umiddelbart gi nøyaktige data om nåværende saldoer, utført arbeid og kommende leveranser, hvis dataene er forberedt på forhånd og samlet i et datalager der det ikke er behov for å vandre gjennom mapper, åpne dusinvise filer og sammenligne verdier manuelt.

I lang tid brukte byggefirmaer PDF -dokumenter, DWG -tegninger, RVT -modeller og hundrevis og tusenvis av Excel -tabeller og andre ulike formater, som lagres i bestemte mapper på selskapets servere, noe som

gjør det vanskelig å søke etter informasjon, kontrollere og analysere den. Resultatet er at filene som blir laggende igjen etter at et prosjekt er avsluttet, som oftest flyttes tilbake til serveren i arkivmapper som i praksis ikke blir brukt i fremtiden. Slik tradisjonell filbasert datalagring mister sin relevans etter hvert som dataflyten øker, fordi den er sårbar for menneskelige feil.

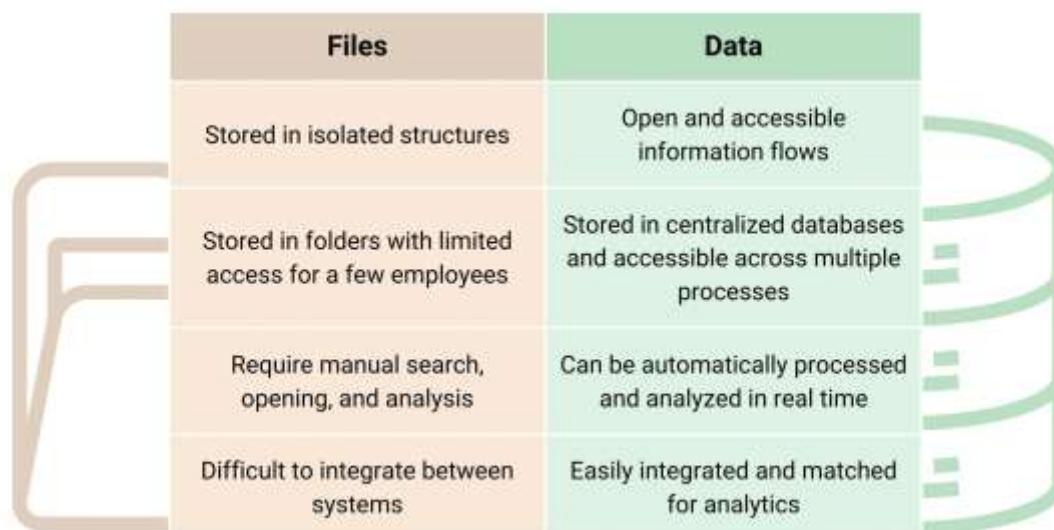
En fil er bare en isolert beholder der data lagres. Filer er laget for mennesker, ikke for systemer, og de krever derfor manuell åpning, lesing og tolkning. Eksempler på dette er en Excel -tabell, et PDF -dokument eller en CAD -tegning som må åpnes spesifikt i et bestemt verktøy for å få tilgang til den ønskede informasjonen. Uten strukturert gjenfinning og behandling forblir informasjonen i dokumentet ubrukt.

Data er på sin side maskinlesbar informasjon som kobles, oppdateres og analyseres automatisk. I et enkelt datavarehus (f.eks. database, DWH eller Data Lake) er informasjonen representert i form av tabeller, poster og relasjoner. Dette muliggjør enhetlig lagring, automatiserte spørrenganger, verdianalyser og rapportering i sanntid.

Ved å bruke data i stedet for filer (fig. 8.1-1) elimineres den manuelle søkerprosessen, og behandlingen blir mer enhetlig. Bedrifter som allerede har implementert en slik tilnærming, får et konkurransefortrinn på grunn av den raske tilgangen til informasjon og muligheten til raskt å integrere den i forretningsprosessene.

Overgangen fra filer til data er en uunngåelig endring som vil forme byggebransjens fremtid.

Alle selskaper i byggebransjen vil stå overfor et viktig valg: Fortsette å lagre informasjon i ulike filer og siloer som må leses av mennesker ved hjelp av spesialprogrammer, eller omdanne den i de første stadiene av behandlingen til strukturerte data og skape ett enkelt integrert digitalt grunnlag for automatisert prosjektstyring.



Files	Data
Stored in isolated structures	Open and accessible information flows
Stored in folders with limited access for a few employees	Stored in centralized databases and accessible across multiple processes
Require manual search, opening, and analysis	Can be automatically processed and analyzed in real time
Difficult to integrate between systems	Easily integrated and matched for analytics

Figur 8.1-1 Utviklingen av informasjonsflyt: fra isolerte filer til integrerte data.

I takt med informasjonseksplosjonen blir tradisjonelle metoder for lagring og behandling av filer stadig mindre effektive. I byggebransjen, som i andre sektorer, er det ikke lenger tilstrekkelig å basere seg på ulike filmapper med forskjellige filformater eller databaser som ikke er koblet sammen.

Bedrifter som ønsker å forbli konkurransedyktige i den digitale tidsalderen, vil unngåelig gå over til integrerte digitale plattformer, ta i bruk stordataknologi og automatiserte analysesystemer.

Overgangen fra filbasert lagring til datadrevne arbeidsflyter vil kreve nytenkning rundt informasjonshåndtering og et bevisst valg av formater som egner seg for videre integrering i sentraliserte repositorier. Dette valget vil avgjøre hvor effektivt data kan behandles, hvor raskt man kan få tilgang til dem og hvor enkelt de kan integreres i bedriftens digitale prosesser.

Lagring av stordata: Analyse av populære formater og deres effektivitet

Lagringsformater spiller en nøkkelrolle når det gjelder skalerbarhet, pålitelighet og ytelse i analyseinfrastrukturen. For å analysere og behandle data - for eksempel filtrering, gruppering og aggregering - brukte eksemplene våre Pandas DataFrame - en populær struktur for å arbeide med data i RAM.

Pandas DataFrame har imidlertid ikke sitt eget lagringsformat, så når behandlingen er fullført, eksporteres dataene til et av de eksterne formatene - som oftest CSV eller XLSX. Disse tabellformatene er enkle å utveksle og kompatible med de fleste eksterne systemer, men har en rekke begrensninger: lav lagringseffektivitet, manglende komprimering og dårlig støtte for versjonering:

- **CSV** (Comma-Separated Values): et enkelt tekstformat som støttes av en rekke plattformer og verktøy. Det er enkelt å bruke, men støtter ikke komplekse datatyper og komprimering.
- **XLSX** (Excel Open XML Spreadsheet): et Microsoft Excel-filformat som støtter sofistikerte funksjoner som formler, diagrammer og styling. Det er nyttig for manuell dataanalyse og visualisering, men er ikke optimalisert for databehandling i stor skala.

I tillegg til de populære tabellformatene XLSX og CSV finnes det flere populære formater for effektiv lagring av strukturerte data (figur 8.1-2), som alle har unike fordeler avhengig av spesifikke krav til datalagring og analyse:

- **Apache Parquet**: et filformat for kolonnelagring av data som er optimalisert for bruk i dataanalysesystemer. Det tilbyr effektive datakomprimerings- og kodingsordninger, noe som gjør det ideelt for komplekse datastrukturer og behandling av store datamengder.
- **Apache ORC** (Optimised Row Columnar): i likhet med Parquet gir ORC høy komprimering og effektiv datalagring. Den er optimalisert for tunge leseoperasjoner og egner seg godt for lagring av datasjøer.
- **JSON** (JavaScript Object Notation): selv om JSON ikke er like effektivt når det gjelder datalagring sammenlignet med binære formater som Parquet eller ORC, er det svært tilgjengelig og enkelt å jobbe med, noe som gjør det ideelt for skript der lesbarhet og nettkompatibilitet er viktig.
- **Feather**: et raskt, lett og brukervennlig analytisk orientert binært kolonnelagringsformat for data. Det er utviklet for effektiv overføring av data mellom Python (Pandas) og R, noe som gjør det til et utmerket valg for prosjekter som involverer disse programmeringsmiljøene.
- **HDF5** (Hierarchical Data Format versjon 5): utviklet for lagring og organisering av store datamengder. Det støtter et bredt spekter av datatyper og egner seg godt for arbeid med komplekse datasamlinger. HDF5 er spesielt populært innen vitenskapelig databehandling på grunn av sin evne

til å lagre og få tilgang til store datasett på en effektiv måte.

		XLSX	CSV	Apache Parquet	HDFS	Pandas DataFrame
	Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
	Compression	Built-in	None	High	Built-in	None (in-memory)
	Performance	Low	Medium	High	High	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Medium	Medium	Low
	Data Type Support	Limited	Very limited	Extended	Extended	Extended
	Scalability	Low	Low	High	High	Medium (memory limited)

Fig. 8.1-2 Sammenligning av dataformater som viser de viktigste forskjellene når det gjelder lagring og behandling.

For å kunne utføre en sammenlignende analyse av formatene som ble brukt i Load-fasen av ETL-prosessen, ble det laget en tabell som viser filstørrelser og lesetider (fig. 8.1-3). Filer med identiske data ble brukt i studien: Tabellen inneholdt 10 000 rader og 10 kolonner fylt med tilfeldige verdier.

Følgende lagringsformater er inkludert i studien: CSV, Parquet, XLSX og HDF5, samt deres komprimerte versjoner i ZIP-arkiver. Rådataene ble generert ved hjelp av NumPy-biblioteket og representert som en Pandas DataFrame-struktur. Testprosessen besto av følgende trinn:

- **Fillagring:** Datarammen lagres i fire forskjellige formater: CSV, Parquet, XLSX og HDF5. Hvert format har unike egenskaper i måten det lagrer data på, noe som påvirker filstørrelsen og lesehastigheten.
- **ZIP-filkomprimering:** For å analysere effektiviteten av standard komprimering ble hver fil ytterligere komprimert til et ZIP-arkiv.
- **Fillesing (ETL - Load):** Lesetiden ble målt for hver fil etter at den var pakket ut fra ZIP-filen. Dette gjør det mulig å estimere hastigheten på datatilgangen etter utpakking fra arkivet.

Det er viktig å merke seg at Pandas DataFrame ikke ble brukt direkte i analysen av størrelse eller lesetid, ettersom det ikke representerer et lagringsformat i seg selv. Den fungerte kun som en mellomstruktur for generering og påfølgende lagring av data i ulike formater.



Fig. 8.1-3 Sammenligning av lagringsformater etter størrelse og lesehastighet.

CSV og HDF5 filer (figur 8.1-3) har høy komprimeringseffektivitet, noe som reduserer størrelsen betydelig når de pakkes i ZIP, noe som kan være spesielt nyttig i scenarier som krever lagringsoptimalisering. XLSX - filer er derimot praktisk talt ikke komprimerbare, og størrelsen i ZIP forblir sammenlignbar med originalen, noe som gjør dem mindre egnet til bruk i store datamengder eller i miljøer der rask datatilgang er viktig. I tillegg er lesetiden for XLSX betydelig høyere enn for andre formater, noe som gjør det mindre egnet for raske dataoperasjoner. Apache Parquet har vist seg å ha høy ytelse for analytiske oppgaver og store datamengder på grunn av sin kolonnestruktur.

Optimaliser lagring med Apache Parquet

Et av de populære formatene for lagring og behandling av stordata er Apache Parquet. Dette formatet er utviklet spesielt for kolonnelagring (i likhet med Pandas), noe som gjør det mulig å redusere minneavtrykket betydelig og øke hastigheten på analytiske spørrslinger. I motsetning til tradisjonelle formater som CSV og XLSX, støtter Parquet innebygd komprimering og er optimalisert for stordatasystemer, inkludert Spark, Hadoop og skylagring.

Blant de viktigste funksjonene i Parquet er støtte for datakomprimering og -koding, noe som reduserer lagningsstørrelsen betydelig og gjør det raskere å lese data ved å arbeide direkte på de ønskede kolonnene i stedet for på alle radene med data.

For å illustrere hvor enkelt det er å få den nødvendige koden for å konvertere data til Apache Parquet, la oss bruke LLM.

- ✉ Send en tekstforespørsel til LLM chat (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Skriv kode for å lagre data fra Pandas DataFrame til Apache Parquet. ↵

- ✉ LLMs svar :

The screenshot shows a LLM chat interface with several icons at the top: Create in LLM Chat, a large blue circle with a white question mark, and icons for CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, and QWEN. Below this is a code editor window titled 'Pandas to parquet.py'. The code is as follows:

```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

At the bottom of the code editor is a toolbar with icons for Run, Stop, PC, and a save symbol, along with a 'Run in IDE' button.

Figur 8.1-4 Overføring av dataframe-data fra RAM til et lagringseffektivt Apache Parquet-format ved hjelp av noen få linjer med Python .

Neste eksempel: La oss simulere ETL-prosessen med data lagret i formatet Parquet for å filtrere prosjekter etter en bestemt verdi på ett av attributtene "cost_million_usd" (fig. 8.1-4).

- ✉ Mens chatten fortsetter, sender du en tekstforespørsel til LLM:

Skriv kode der vi ønsker å filtrere dataene i en tabell og bare lagre de prosjektene (tabellradene) fra Apache Parquet-dataene, hvis kostnad (parameter cost_million_usd) overstiger 150 millioner dollar. ↵

LLMs svar :

```

1 # Loading construction project data
2 df_loaded_projects = pd.read_parquet(parquet_path_projects)
3
4 # Filtering out projects with a value greater than $150 million
5 df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]
6
7 # Saving the filtered data to a new Parquet file
8 filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
9 df_filtered_projects.to_parquet(filtered_parquet_path_projects)
10
11 print("Filtered construction projects data saved to Lakehouse.")

```

Run in IDE

Figur 8.1-5 ETL-prosessen når du arbeider med data i Apache Parquet -formatet, ser likedan ut som med andre strukturerte formater.

Ved å bruke Parquet-formatet (i motsetning til XLSX, CSV osv.) reduserer mengden informasjon som lagres, og gjør søkeoperasjoner raskere. Dette gjør det utmerket til både lagring og analyse av data. Parquet kan integreres med ulike prosesseringssystemer, noe som gir effektiv tilgang i hybridarkitekturen.

Et effektivt lagringsformat er imidlertid bare ett element i en komplett dataopplevelse. For å skape et bærekraftig og skalerbart miljø er det nødvendig med en veldesignet arkitektur for datahåndtering. Systemer i DWH-klassen (Data Warehouse) oppfyller denne funksjonen. De sørger for aggregering av data fra heterogene kilder, innsyn i forretningsprosesser og mulighet for komplekse analyser ved hjelp av BI-verktøy og maskinlæringsalgoritmer.

DWH: Data Warehouse datavarehus

På samme måte som formatet Parquet er optimalisert for effektiv lagring av store mengder informasjon, er Data Warehouse optimalisert for integrering og strukturering av data for å støtte analyser, prognosering og beslutningstaking.

I dagens bedrifter kommer data fra mange ulike kilder: ERP-, CAFM-, CPM-, CRM-systemer, regnskaps- og lagerstyring, digitale CAD-modeller av bygninger, IoT-sensorer og andre løsninger. For å få et helhetlig bilde er det ikke nok å bare samle inn data - de må organiseres, standardiseres og sentraliseres i ett enkelt datalager. Det er nettopp dette DWH gjør - et centralisert lagringssystem som gjør det mulig å samle informasjon fra ulike kilder, strukturere den og gjøre den tilgjengelig for analyse og strategisk ledelse.

DWH (Data Warehouse) er et centralisert datavarehussystem som samler informasjon fra flere kilder, strukturerer den og gjør den tilgjengelig for analyse og rapportering.

I mange bedrifter er dataene spredt på tvers av ulike systemer, noe vi diskuterte i de første delene av boken (figur 1.2-4). DWH integrerer disse kildene og sikrer full åpenhet og pålitelighet i informasjonen. Et DWH-datavarehus er en spesialisert database (en stor database) som samler inn, behandler og lagrer data fra flere kilder. De viktigste kjennetegnene ved et DWH er

- **Ved hjelp av ETL -prosesser** Extract, Transform, Load) - trekke ut data fra kilder, rense dem, transformere dem, laste dem inn i repositoriet og automatisere disse prosessene, som ble diskutert i del syv av boken.
- **Datagrularitet** - data i DWH kan lagres både i aggregert form (sammendragsrapporter) og i granulær form (rådata). Fra og med 2024 er det CAD- leverandører som har begynt å snakke om granulære data [125], noe som kanskje indikerer at bransjen forbereder seg på overgangen til spesialisert skylagring for håndtering av digitale bygningsmodelldata.
- **Støtte for analyse og prediktiv analyse** - datavarehus danner grunnlaget for BI-verktøy, Big Data - analyse og maskinlæring.

DWH danner grunnlaget for business intelligence, som muliggjør analyse av nøkkelindikatorer, prognosenter for salg, innkjøp og kostnader, samt automatisert rapportering og datavisualisering (figur 8.1-6).

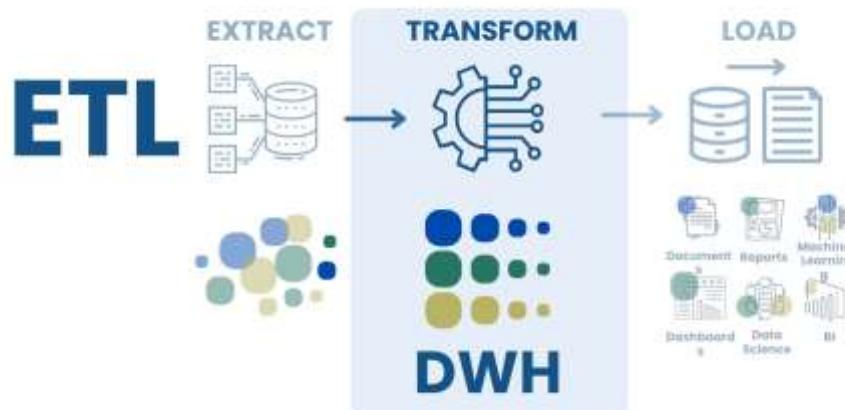


Fig. 8.1-6 I en ETL -prosess kan DWH fungere som et sentralt lager der data som hentes ut fra ulike systemer, gjennomgår transformasjons- og utlastingsfaser.

DWH spiller en nøkkelrolle når det gjelder å integrere, rense og strukturere informasjon, og danner et solid grunnlag for business intelligence og beslutningsprosesser. Men i dagens miljø, der datamengdene vokser raskt og datakildene blir stadig mer mangfoldige, må den tradisjonelle DWH-tilnærmingen til informasjonslagring ofte utvides i form av ELT og Data Lake

Data Lake - utviklingen fra ETL til ELT: fra tradisjonellrensing til fleksibel behandling

Klassiske DWH - datavarehus, som er utformet for å lagre strukturerte data i et format som er optimalisert for analytiske spørrslinger, har hatt sine begrensninger når det gjelder håndtering av ustukturerte data og skalerbarhet. Som svar på disse utfordringene har det vokst frem Data Lakes), som tilbyr fleksibel lagring

av store mengder heterogene data.

Data Lake tilbyr en alternativ DWH -tilnærming som gjør det mulig å arbeide med ustrukturerte, halvstrukturerte og ubearbeidede data uten et fast skjema på forhånd. Denne lagringsmetoden er ofte relevant for databehandling i sanntid, maskinlæring og avansert analyse. I motsetning til DWH, som strukturerer og aggregerer data før de lastes inn, gjør Data Lake det mulig å lagre informasjon i sin rå form, noe som gir fleksibilitet og skalerbarhet

Det var frustrasjonen over tradisjonelle datavarehus (RDBMS, DWH) og interessen for "big data" som førte til fremveksten av datasjøer, der data i stedet for kompleks ETL nå ganske enkelt lastes opp til et løst strukturert datalager, og behandlingen skjer i analysestadiet:

- I tradisjonelle datavarehus blir dataene vanligvis forhåndsbehandlet, transformert og renset (ETL - Extract, Transform, Load) før de lastes inn i datavarehuset (figur 8.1-6). Dette betyr at dataene struktureres og optimaliseres for spesifikke fremtidige analyser og rapporteringsoppgaver. Det legges vekt på å opprettholde høy spørtingsytelse og dataintegritet. Denne tilnærmingen kan imidlertid være kostbar og mindre fleksibel når det gjelder å integrere nye datatyper og dataskjemaer som endres raskt.
- Datasjøer, derimot, er utviklet for å lagre store mengder rådata i sitt opprinnelige format (figur 8.1-7). ETL-prosessen (Extract, Transform, Load), er i ferd med å bli erstattet av ELT (Extract, Load, Transform), der data først lastes inn i lageret "som de er", og først deretter kan transformeres og analyseres etter behov. Dette gir større fleksibilitet og mulighet til å lagre heterogene data, inkludert ustrukturerte data som tekst, bilder og logger.

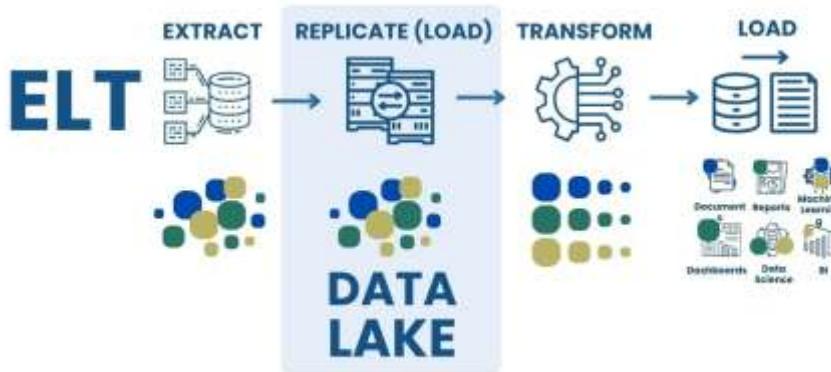


Fig. 8.1-7 I motsetning til ETL bruker Data Lake ELT, der informasjonen først lastes opp i "rå" form, og transformasjonen utføres i opplastingsfasen.

Tradisjonelle datavarehus fokuserer på forhåndsbehandling av data for å sikre høy spørtingsytelse, mens datasjøer prioritiserer fleksibilitet: De lagrer rådata og transformerer dem etter behov (figur 8.1-8).

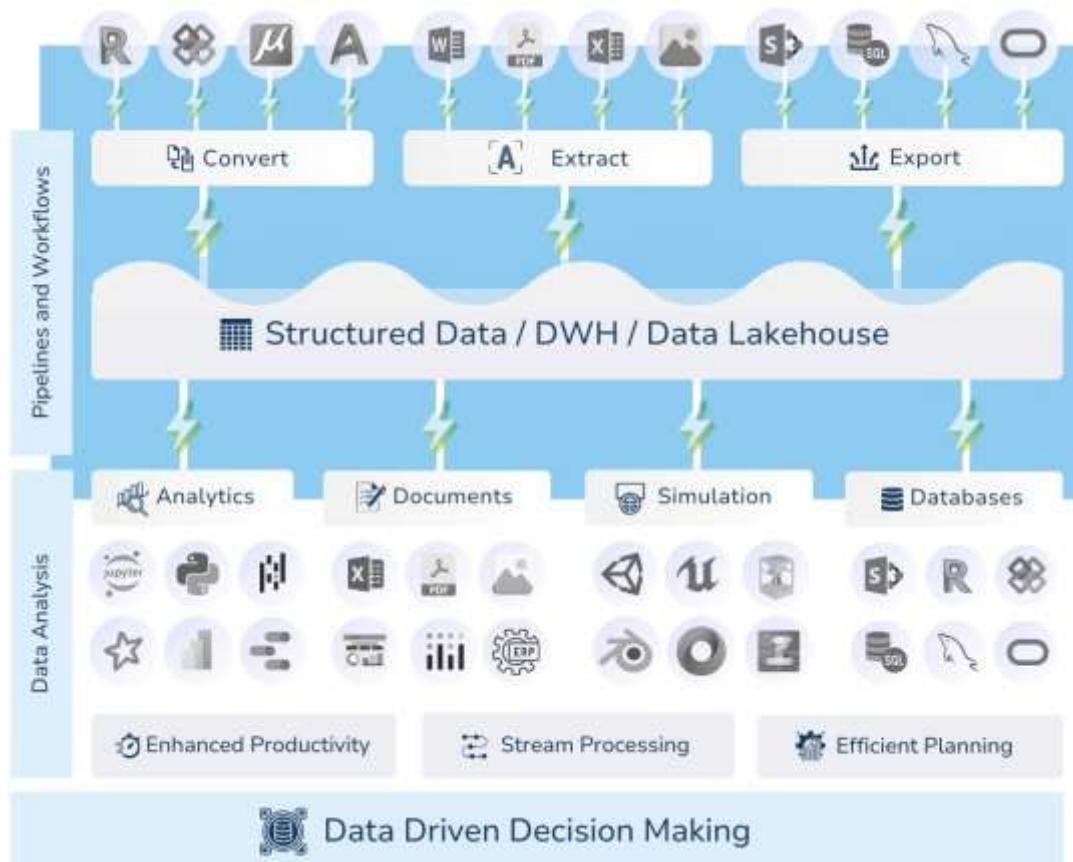


Fig. 8.1-8 Moderne lagringskonsepter tar sikte på å lagre og behandle alle typer data for beslutningsformål.

Til tross for alle fordelene er datasjøer imidlertid ikke uten ulemper. Mangelen på en streng struktur og kompleksiteten i informasjonshåndteringen kan føre til kaos, der data dupliseres, motsier hverandre eller blir irrelevante. I tillegg krever det mye arbeid å søke etter og analysere data i et slikt arkiv, spesielt når man har å gjøre med heterogen informasjon. For å overvinne disse begrensningene og kombinere de beste egenskapene til tradisjonelle datavarehus og datasjøer, ble Data Lakehouse-arkitekturen utviklet.

Data Lakehouse-arkitektur: synergi mellom datalagre og datasjøer

For å kombinere de beste egenskapene til DWH (strukturert, håndterbar, høy ytelsesanalyse) og Data Lake (skalerbarhet, håndtering av heterogene data), ble Data Lakehouse-tilnærmingen utviklet. Denne arkitekturen kombinerer fleksibiliteten til datasjøer med de kraftige behandlings- og administrasjonsverktøyene som er typiske for tradisjonelle datalagre, og skaper en balanse mellom lagring, analyse og maskinlæring. Data Lakehouse er en syntese av datasjøer og datavarehus, som kombinerer fleksibiliteten og skalerbarheten til førstnevnte med håndterbarheten og spørringsoptimaliseringen til sistnevnte.

Data Lakehouse er en arkitektonisk tilnærming som søker å kombinere fleksibiliteten og skalerbarheten til datasjøer med administrerbarheten og spørringssyntesen til datavarehus (figur 8.1-9).

Viktige funksjoner i Data Lakehouse inkluderer

- **Åpent datalagringsformat:** Bruk av åpne formater for datalagring, for eksempel Apache Parquet, gir effektivitet og optimaliserte spørninger.
- **Skrivebeskyttet skjema:** I motsetning til den tradisjonelle tilnærmingen med et skrivebeskyttet skjema i DWH, støtter Lakehouse et skrivebeskyttet skjema, noe som gir større fleksibilitet i håndteringen av datastrukturen.
- **Fleksibel og skalerbar:** Støtter lagring og analyse av strukturerte og ustrukturerte data, og gir høy spørringssyntese gjennom optimalisering på lagringsnivå.

Data Lakehouse tilbyr en kompromissløsning som kombinerer fordelene med begge tilnærmingene, noe som gjør den ideell for moderne analysearbeid som krever fleksibilitet i databehandlingen.



Fig. 8.1-9 Data Lakehouse er neste generasjon lagringssystemer som er utviklet for å møte komplekse og stadig skiftende krav.

Tanken bak moderne datavarehus virker enkel: Hvis alle dataene er samlet på ett sted, er det lettere å analysere dem. Men i praksis er ikke alt så enkelt. Tenk deg at et selskap bestemmer seg for å gå helt bort fra de vanlige regnskaps- og styringssystemene (ERP, PMIS, CAFM eller andre), og erstatte dem med en stor datasjø som alle har tilgang til. Hva vil da skje? Sannsynligvis kaos: Data vil bli duplisert, motstridende og kritisk informasjon vil gå tapt eller bli ødelagt. Selv om datasjøen bare brukes til analyse, vil den bli alvorlig kompromittert uten riktig styring:

- Data er vanskelig å forstå: I konvensjonelle systemer har data en klar struktur, men i en innsjø er det bare en enorm ansamling av filer og tabeller. For å finne noe, må spesialist finne ut hva hver rad og kolonne er ansvarlig for.

- Data kan være unøyaktige: Hvis mange versjoner av samme informasjon er lagret på ett sted, er det vanskelig å vite hvilken versjon som er oppdatert. Resultatet er at beslutninger blir tatt på grunnlag av utdaterte eller feilaktige data.
- Det er vanskelig å forberede data for arbeid: dataene må ikke bare lagres, men også presenteres i en praktisk form - i form av rapporter, grafer, tabeller. I tradisjonelle systemer gjøres dette automatisk, men i datasjøer krever det ytterligere behandling.

Derfor har hvert datavarehuskonsept sine egne egenskaper, behandlingsmetoder og forretningsapplikasjoner. Tradisjonelle databaser fokuserer på transaksjonsoperasjoner, datavarehus (DWH) gir en struktur for analyse, datasjøer (Data Lake) lagrer informasjon i rå form, og hybridlagre (Data Lakehouse) kombinerer fordelene med DWH og Data Lake (figur 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Fig. 8.1-10 DWH, Data Lake og Data Lakehouse: viktige forskjeller i datatyper, bruksscenerier, behandlingsmetoder og lagringsmetoder.

Valg av lagringsarkitektur er en kompleks prosess som avhenger av forretningsbehov, informasjonsvolum og analysekrav. Hver løsning har sine fordeler og ulemper: DWH gir struktur, Data Lake gir fleksibilitet, og Lakehouse gir en balanse mellom de to. Organisasjoner er sjeldent begrenset til én enkelt dataarkitektur.

Uansett hvilken arkitektur som velges, er automatiserte datahåndteringssystemer betydelig bedre enn manuelle metoder. De minimerer menneskelige feil, gjør informasjonsbehandlingen raskere og sikrer gjennomsiktighet og sporbarhet i alle ledd av forretningsprosessene.

Og mens sentraliserte datavarehus allerede har blitt en bransjestandard i mange deler av økonomien, er situasjonen i byggebransjen fortsatt fragmentert. Dataene her er fordelt på ulike plattformer (CDE, PMIS, ERP osv.), noe som gjør det vanskelig å skape et helhetlig bilde av hva som foregår, og det krever arkitekturen som kan kombinere disse kildene til et helhetlig, analytisk anvendelig digitalt miljø.

CDE, PMIS, ERP eller DWH og Data Lake

Noen bygg- og anleggsbedrifter bruker allerede konseptet Common Data Environment (CDE) i henhold til ISO 19650. CDE har i hovedsak de samme funksjonene som et datavarehus (DWH) i andre bransjer:

sentralisering av informasjon, versjonskontroll og tilgang til validert informasjon.

Et Common Data Environment (CDE) er et sentralisert digitalt område som brukes til å administrere, lagre, dele og samarbeide om prosjektinformasjon gjennom alle faser av et anleggs livssyklus. CDE implementeres ofte ved hjelp av skybaserte teknologier og integreres med CAD (BIM) -systemer.

Finans-, detaljhandels-, logistikk- og industrisektoren har i flere tiår brukt sentraliserte datahåndteringssystemer som kombinerer informasjon fra ulike kilder, kontrollerer relevansen og tilbyr analyser. CDE tar disse prinsippene videre ved å tilpasse dem til utfordringene innen bygningsdesign og livssyklusstyring.

I likhet med DWH strukturerer CDE data, fanger opp endringer og gir tilgang til verifisert informasjon fra ett enkelt sted. Med overgangen til skyen og integrering med analyseverktøy blir forskjellene mellom de to stadig mindre tydelige. Hvis man legger til CDE granulære data, et konsept som har vært diskutert av CAD -leverandører siden 2023[93, 125], kan man se enda flere paralleller med klassisk DWH.

Tidligere i kapittelet "Construction ERP og PMIS systems" har vi allerede sett på PMIS (Project Management Information System) og ERP (Enterprise Resource Planning). I byggeprosjekter samarbeider CDE og PMIS: CDE fungerer som et datalager for blant annet tegninger, modeller og prosjektdokumentasjon, mens PMIS håndterer prosesser som kontroll av tidsfrister, oppgaver, ressurser og budsjetter.

ERP, som er ansvarlig for å styre virksomheten som helhet (økonomi, innkjøp, personal, produksjon), kan integreres med PMIS, som gir kostnads- og budsjettkontroll på bedriftsnivå. For analyse og rapportering kan DWH brukes til å samle inn, strukturere og aggregere data fra CDE, PMIS og ERP for å evaluere økonomiske KPI-er (ROI) og identifisere mønstre. Data Lake (DL) kan på sin side utfylle DWH ved å lagre rådata og ustrukturerte data (f.eks. logger, sensordata og bilder). Disse dataene kan behandles og lastes inn i DWH for videre analyse.

CDE og PMIS fokuserer på prosjektledelse, ERP fokuserer på forretningsprosesser, og DWH og Data Lake fokuserer på analyse og datadrift.

Når man sammenligner CDE, PMIS og ERP med DWH og Data Lake, ser man betydelige forskjeller når det gjelder leverandøruavhengighet, kostnader, integrasjonsfleksibilitet, datauavhengighet, hvor raskt man kan tilpasse seg endringer og analytiske muligheter (figur 8.1-11). Tradisjonelle systemer som CDE, PMIS og ERP er ofte bundet til spesifikke leverandørløsninger og standarder, noe som gjør dem mindre fleksible og øker kostnadene på grunn av lisenser og support. I tillegg er dataene i slike systemer ofte innkapslet i proprietære, lukkede formater, noe som begrenser bruken og analysene av dem.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Fig. 8.1-11 DWH og Data Lake tilbyr større fleksibilitet og dataavhengighet enn systemer som CDE, PMIS og ERP.

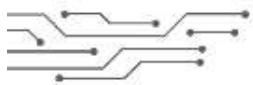
DWH og Data Lake gir derimot større fleksibilitet når det gjelder integrering med ulike datakilder, og bruken av åpne teknologier og plattformer bidrar til å redusere de totale eierkostnadene. DWH og Data Lake støtter dessuten et bredt spekter av analyseverktøy, noe som forbedrer analyse- og administrasjonsmulighetene.

Med utviklingen av verktøy for omvendt konstruksjon av DAK-formater og tilgang til DAK-applikasjonsdatabaser blir spørsmålet mer og mer akutt: Hvor berettiget er det å fortsette å bruke lukkede, isolerte plattformer hvis konstruksjonsdataene skal være tilgjengelige for et bredt spekter av spesialister som arbeider i dusinvis av entreprenører og konstruksjonsorganisasjoner?

Denne avhengigheten av leverandørspesifikk teknologi kan begrense fleksibiliteten i datahåndtering, forsinke responsen på prosjektendringer og hindre effektivt samarbeid mellom deltakerne.

Tradisjonelle tilnærminger til datahåndtering - inkludert DWH, Data Lake, CDE og PMIS - har først og fremst fokusert på lagring, strukturering og behandling av informasjon. Med utviklingen av kunstig intelligens og maskinlæring er det imidlertid et økende behov for nye måter å organisere data på, som ikke bare aggregerer, men også avdekker komplekse sammenhenger, finner skjulte mønstre og gir umiddelbar tilgang til den mest relevante informasjonen.

Vektordatabaser - en ny type lagring som er optimalisert for høydimensjonale innstøpninger - er i ferd med å spille en spesiell rolle i denne retningen.



KAPITTEL 8.2.

DATAVAREHUSADMINISTRASJON OG FOREBYGGING AV KAOS

Vektordatabaser og den avgrensende boksen

Vektordatabaser er en ny type databaser som ikke bare lagrer data, men som også gjør det mulig å søke etter mening, sammenligne objekter ut fra semantisk nærhet og skape intelligente systemer: fra anbefalinger til automatisk analyse og generering av kontekst. I motsetning til tradisjonelle databaser som fokuserer på eksakte treff, finner vektordatabaser lignende objekter basert på attributter - selv om ikke finnes noe eksakt treff.

En vektordatabase er en spesialisert type database som lagrer data som flerdimensjonale vektorer, der hver vektor representerer bestemte egenskaper eller kvaliteter. Vektorene kan ha ulikt antall dimensjoner, avhengig av hvor komplekse dataene er (i ett tilfelle kan det dreie seg om noen få dimensjoner, i et annet om flere tusen).

Den største fordelen med vektordatabaser er at man kan søke etter semantisk relevans i stedet for nøyaktig samsvar mellom verdier. I stedet for SQL- og Pandas -forespørsler med "equals"- eller "contains"-filtre, brukes søk etter nærmeste naboer (k-NN) (vi skal snakke mer om k-NN i neste del av boken) i funksjonsrommet.

Med utviklingen av LLM (Large Language Models) og generative modeller er interaksjonen med databaser i ferd med å endre seg. Det er nå mulig å stille spørsmål om data på naturlig språk, gjøre semantiske søk i dokumenter, automatisk trekke ut nøkkelbegreper og bygge kontekstuelle relasjoner mellom objekter - alt uten at man trenger å beherske SQL eller ha kjennskap til tabellstrukturen. Dette ble diskutert mer inngående i avsnittet "LLM-er og deres rolle i databehandling og forretningsprosesser".

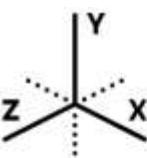
Det er imidlertid viktig å være klar over at LLM-er ikke automatisk strukturerer og ordner informasjon. Modellen flyter bare gjennom dataene og finner de mest relevante dataene basert på konteksten for spørringen. Hvis dataene ikke er renset eller transformert på forhånd, vil et dypt søk være som å prøve å finne et svar i digitalt "søppel" - det kan fungere, men kvaliteten på resultatene vil være lavere. Det ideelle er om dataene kan struktureres (f.eks. ved å oversette dokumenter til Markdown) og lastes inn i en vektordatabase. Dette øker nøyaktigheten og relevansen av resultatet betydelig.

Opprinnelig ble vektordatabaser brukt i maskinlæring, men i dag finner de stadig flere bruksområder utenfor maskinlæring - i søkemotorer, personalisering av innhold og intelligente analyser.

Et av de mest åpenbare eksemplene på vektortilnærmingen i konstruksjonen er Bounding Box (avgrensende parallellpipedum). Det er en geometrisk konstruksjon som beskriver grensene for et objekt i det tredimensjonale rommet. Bounding Box defineres av minimums- og maksimumskoordinatene for X, Y og Z, som danner en "boks" rundt objektet. Denne metoden gjør det mulig å estimere størrelsen og plasseringen av et element uten å måtte analysere hele geometrien.

Hver Bounding Box kan representeres som en vektor i et flerdimensjonalt rom: for eksempel [x, y, z, bredde,

høyde, dybde] - allerede 6 dimensjoner (Fig. 8.2-1).



Bounding Box

	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Fig. 8.2-1 Bounding Box -elementenes koordinatinformasjon og deres plassering i prosjektmodellen er analog med en vektordatabase.

Denne datarepresentasjonen gjør det enklere å utføre mange oppgaver, blant annet å se etter skjæringspunkter mellom objekter, planlegge den romlige fordelingen av bygningselementer og utføre automatiserte beregninger. Bounding Box kan fungere som en bro mellom komplekse 3D-modeller og tradisjonelle vektordatabaser, slik at du effektivt kan utnytte fordelene begge tilnærmingene i arkitektonisk og teknisk modellering.

Bounding Box er "vektorisering av geometri", og embeddingmåte å transformere noe abstrakt på) er "vektorisering av mening". Begge tilnærmingene gjør det mulig å gå fra manuelt søk til intelligent søk, enten det dreier seg om 3D -objekter i en prosjektmodell eller begreper i en tekst.

Søk etter objekter i prosjektet (for eksempel "finn alle vinduer med bredde > 1,5 m") ligner på søk etter nærmeste naboer (k-NN) i en vektordatabase, der kriteriene definerer en "sone" i funksjonsrommet. (Vi kommer til å snakke mer om k-NN-søk etter nærmeste naboer i neste del om maskinlæring) (fig. 8.2-2). Hvis vi legger til flere parametere (materiale, vekt, produksjonstid) til attributtene i den avgrensende boksen, blir tabellen til en høydimensjonal vektor, der hvert attributt er en ny dimensjon. Dette ligger nærmere moderne vektorbaserte dimensjonene telles i hundrevis eller tusenvis (f.eks. innbygging fra nevrale nettverk).

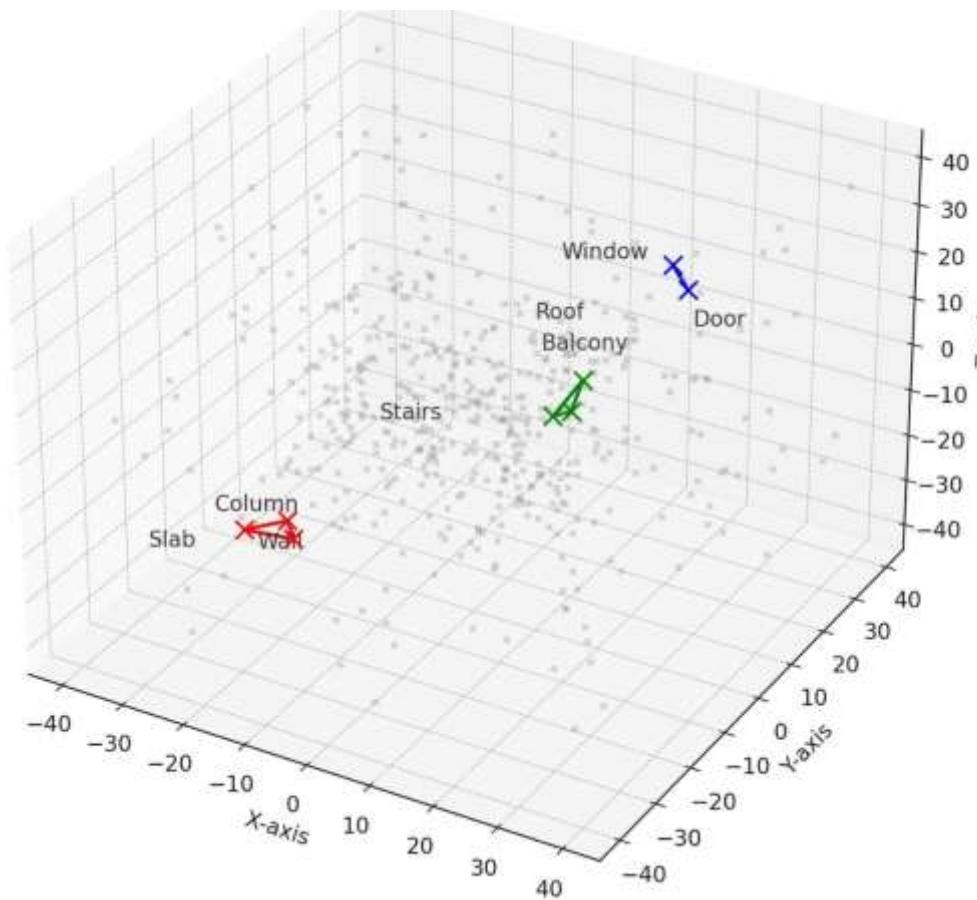


Fig. 8.2-2 Søk etter objekter i prosjektet ved hjelp av vektordatabaser.

Tilnærmingen som brukes i Bounding Box, kan ikke bare brukes på geometriske objekter, men også på tekst- og språkanalyse. Vektorrepresentasjoner av data brukes allerede aktivt i naturlig språkbehandling (NLP). På samme måte som objekter i et byggeprosjekt kan gruppere i henhold til deres romlige nærhet (figur 8.2-2), kan ord i en tekst analyseres i henhold til deres semantiske og kontekstuelle nærhet.

Ordene "arkitekt", "konstruksjon" og "design" vil for eksempel stå ved siden av hverandre i vektorrommet fordi de har en lignende betydning. I LLM gjør denne mekanismen det mulig med automatisk kategorisering, uten behov for manuell kategorisering:

- Identifisere emnet i en tekst
- Utfør semantiske søk på innholdet i dokumenter
- Generer automatiske merknader og tekstsammendrag
- Finn synonymer og relaterte termer

Vektordatabaser gjør det mulig å analysere tekst og finne relaterte termer i den på samme måte som Bounding Box hjelper deg med å analysere romlige objekter i 3D-modeller. Bounding Box-eksempelet med prosjektelementer gjør det lettere å forstå at vektorrepresentasjon ikke er et rent "kunstig" konsept fra ML, men en naturlig måte å strukturere data på for å løse anvendte problemer, enten det dreier seg om å søke etter kolonner i et CAD-prosjekt eller semantisk nære bilder i en database.

Spesialister som jobber med databaser, bør være oppmerksomme på vektorlagre. Utbredelsen av slike databaser indikerer en ny fase i databaseutviklingen, der klassiske relasjonssystemer og AI-orienterte teknologier begynner å flette seg sammen og danne fremtidens hybridløsninger.

Brukere som utvikler komplekse og storskala AI-applikasjoner, vil bruke spesialiserte databaser for vektorsøk. Samtidig er det mer sannsynlig at de som bare trenger separate AI-funksjoner for integrering i eksisterende applikasjoner, vil velge innebygde vektorsøkfunksjoner i databasene de allerede bruker (PostgreSQL, Redis).

Selv om systemer som DWH, Data Lake, CDE, PMIS, vektordatabaser og andre tilbyr ulike tilnærminger til datalagring og -håndtering, er det ikke bare arkitekturen som avgjør hvor effektive de er, men også hvor godt dataene i seg selv er organisert og håndtert. Selv ved bruk av moderne løsninger - det være seg vektordatabaser, klassiske relasjonelle DBMS-er eller datalagre av Data Lake-typen - kan mangelen på klare regler for håndtering, strukturering og oppdatering av data føre til de samme problemene som brukere som arbeider med ulike filer og data i flere formater.

Uten datastyring) kan selv de kraftigste løsningene bli kaotiske og ustukturerte, og gjøre datasjøer til datasumper). For å unngå dette må bedriftene ikke bare velge riktig lagringsarkitektur, men også implementere dataminimalisme), tilgangsstyring og strategier for kvalitetskontroll for å gjøre dataene til et effektivt verktøy for beslutningstaking.

Data Governance), Dataminimalisme) og Data Swamp)

Forståelse og implementering av konseptene Data Governance), Dataminimalisme) og forebygging av datasump) er avgjørende for å lykkes med å administrere datavarehus og levere forretningsverdi (figur 8.2-3).

Ifølge en studie fra Gartner (2017) mislykkes 85% av stordataprosjekter, og en av hovedårsakene er utilstrekkelig datakvalitet og datastyring [144].

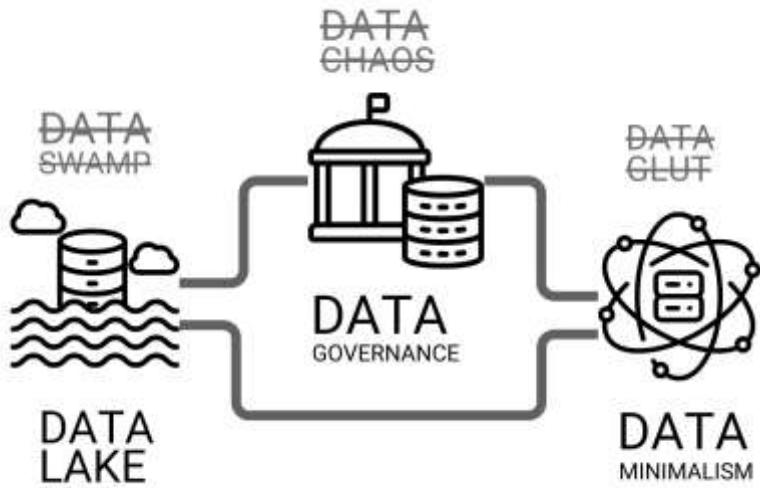


Fig. 8.2-3 Noen av de viktigste aspektene ved datastyring er Data Governance og Dataminimalisme.

Data Governance (Data Governance) er en grunnleggende komponent i datahåndtering, og sikrer at data brukes på en hensiktsmessig og effektiv måte i alle forretningsprosesser. Det handler ikke bare om å etablere regler og prosedyrer, men også om å sørge for at dataene er tilgjengelige, pålitelige og sikre:

- Definere og klassifisere data: En tydelig definisjon og klassifisering av enheter gjør det mulig for organisasjoner å forstå hvilke enheter som trengs i selskapet, og bestemme hvordan de skal brukes.
- Tilgangsrettigheter og administrasjon: Utvikling av retningslinjer og prosedyrer for tilgang til og administrasjon av data sikrer at bare autoriserte brukere får tilgang til bestemte data.
- Beskyttelse av data mot eksterne trusler: Beskyttelse av data mot eksterne trusler er et sentralt aspekt ved datahåndtering. Dette omfatter ikke bare tekniske tiltak, men også opplæring av de ansatte i grunnleggende informasjonssikkerhet.

Dataminimalisme (Data Minimalism) er en tilnærming som går ut på å redusere data til de mest verdifulle og meningsfulle attributtene og enhetene i formasjonen (figur 8.2-4), og dermed redusere kostnadene og forbedre datautnyttelsen:

- Forenkling av beslutningsprosessen: Ved å redusere antallet objekter og deres attributter til de mest relevante, forenkles beslutningsprosessen ved at det kreves mindre tid og ressurser for å analysere og behandle data.
- Fokus på det som er viktig: Ved å velge ut de mest relevante enhetene og attributtene kan du fokusere på den informasjonen som virkelig betyr noe for virksomheten, og eliminere støy og unødvendige data.
- Effektiv ressursallokering: Dataminimering muliggjør mer effektiv ressursallokering, reduserer datalagrings- og behandlingskostnader og forbedrer datakvalitet og -sikkerhet.

Logikken i arbeidet med data bør ikke starte med selve opprettelsen (fig. 8.2-4), men med en forståelse av fremtidige scenarier for bruk av disse dataene allerede før genereringsprosessen starter. Denne tilnærmingen gjør det mulig å definere minimumskravene til attributter, deres typer og grenseverdier på forhånd. Disse kravene danner grunnlaget for å skape korrekte og stabile enheter i informasjonsmodellen.

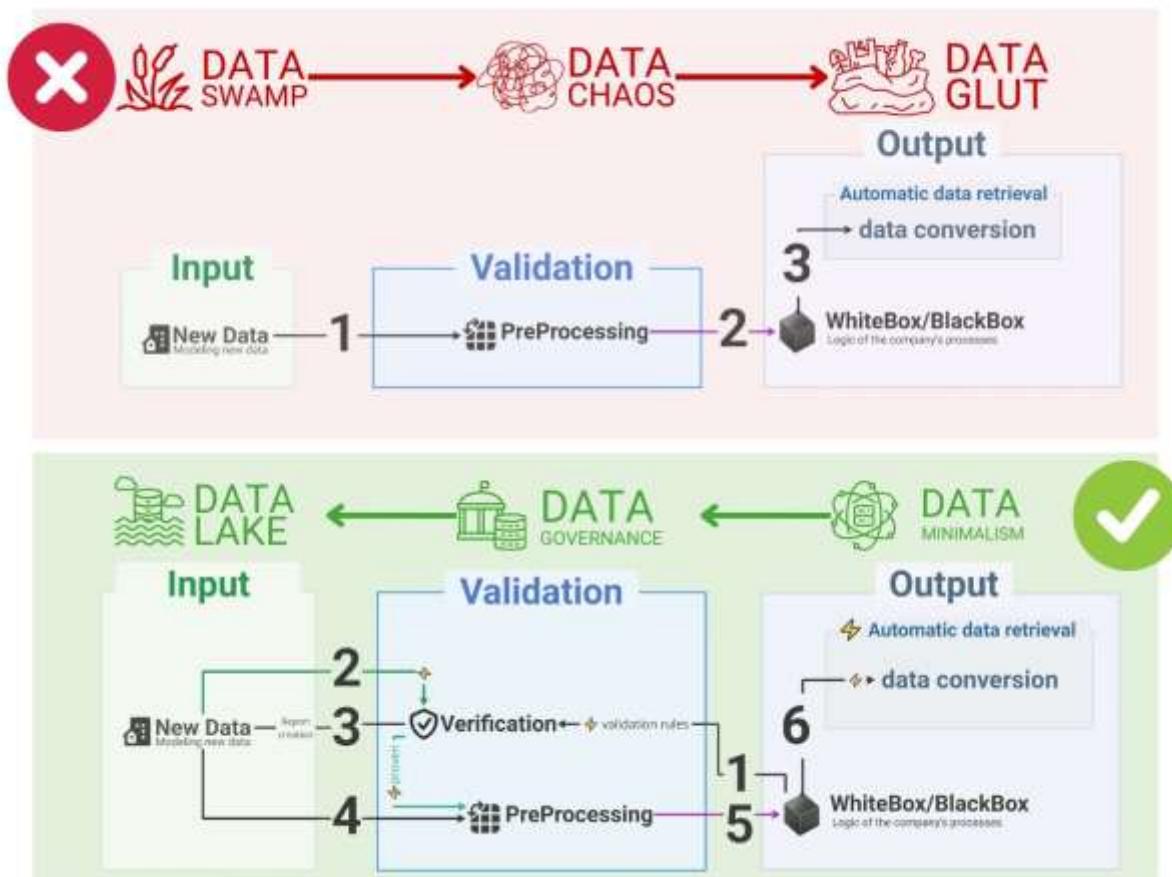
En foreløpig forståelse av formålet med og bruken av dataene bidrar til å skape en struktur som egner seg for analyse. Flere detaljer om tilnæringer til datamodellering på konseptuelt, logisk og fysisk nivå ble diskutert i kapittelet "Datamodellering: konseptuell, logisk og fysisk modell".

I tradisjonelle forretningsprosesser i byggefirmaer ligner databehandling oftere på å dumpe data i en sump, der data først opprettes, og deretter prøver spesialister å integrere dem i andre systemer og verktøy.

Data Swamp (Data Swamp) er resultatet av ukontrollert innsamling og lagring av data uten skikkelig organisering, strukturering og styring, noe som resulterer i ustrukturerte data som er vanskelige å bruke og av liten verdi.

Hvordan unngå at informasjonsflyten blir en hengemyr:

- **Styring av datastruktur:** Ved å sørge for at dataene på er strukturerde og kategoriserte, unngår man at de drukner i data ved å gjøre dem oversiktlige og lett tilgjengelige.
- **Forståelse og tolkning av data:** En tydelig beskrivelse av dataenes opprinnelse, modifikasjoner og betydning sikrer at dataene blir forstått og tolket på riktig måte.
- **Opprettholde datakvaliteten:** Regelmessig vedlikehold og rensing av data bidrar til å opprettholde datakvalitet, relevans og verdi for analyse- og forretningsprosesser.



Figur 8.2-4 For å unngå rot i datavarehuset bør du starte prosessen med å opprette data ved å samle inn krav til attributter.

Ved å integrere prinsippene om datastyring og dataminimalisme i datahåndteringsprosessene og aktivt forhindre at datavarehusene utvikler seg til datasumper, kan organisasjoner maksimere potensialet i dataene sine.

Neste trinn i utviklingen av dataarbeidet, etter å ha løst problemene med forvaltning og minimalisme, er standardisering av automatisk behandling, kvalitetssikring og implementering av metoder som gjør dataene anvendelige for analyse, transformasjon og beslutningstaking. Det er dette som gjøres med DataOps og VectorOps, som er i ferd med å bli viktige verktøy for bedrifter som jobber med stordata og maskinlæring.

DataOps og VectorOps: nye datastandarder

Mens Data Governance er ansvarlig for å kontrollere og organisere data, bidrar DataOps til å sikre at dataene er nøyaktige, konsistente og flyter jevnt i selskapet. Dette er spesielt viktig for en rekke forretningsområder i byggebransjen, der data genereres kontinuerlig og må behandles i tide. DataOps kan for eksempel spille en nøkkelrolle i situasjoner der bygningsinformasjonsmodeller, prosjektkrav og analytiske rapporter må synkroniseres mellom ulike systemer i løpet av én og samme arbeidsdag. Det gjør det mulig å bygge stabile og repeterbare databehandlingsprosesser, noe som reduserer risikoen for forsinkelser og tap av relevant

informasjon.

Datastyring alene er ikke nok - det er avgjørende at data ikke bare lagres, men at de brukes aktivt i den daglige driften. Det er her DataOps - en metodikk som fokuserer på automatisering, integrering og kontinuerlig dataflyt - kommer inn i bildet.

DataOps fokuserer på å forbedre samarbeid, integrering og automatisering av dataflyten i organisasjoner. DataOps-praksiser fremmer nøyaktighet, konsistens og tilgjengelighet av data, noe som er avgjørende for datasentriske applikasjoner.

Viktige verktøy i DataOps økosystem er Apache Airflow (figur 7.4-4) for orkestrering av arbeidsflyter, og Apache NiFi (figur 7.4-5) for ruting og transformering av dataflyter. Sammen muliggjør disse teknologiene fleksible, pålitelige og skalarbare datarørledninger som automatisk behandler, kontrollerer og integrerer informasjon mellom systemer mer informasjon i kapittelet "Automatisk ETL -conveyor"). Når DataOps-tilnærmingen skal implementeres i byggeprosesser, er det viktig å ta hensyn til fire grunnleggende aspekter:

1. **Mennesker og verktøy er viktigere enn data:** siloformede datalagre kan bli sett på som et stort problem, men virkeligheten er mer kompleks. I tillegg til datafragmentering spiller isolasjonen av teamene og de ulike verktøyene de bruker, en viktig rolle. I byggebransjen jobber spesialister fra ulike fagområder med data: dataingeniører og analytikere, BI- og visualiseringsteam, samt prosjektledelse og kvalitetsekspertene. Hver av dem har ulike måter å jobbe på, og det blir derfor viktig å skape et økosystem der dataene flyter fritt mellom deltakerne, slik at man får én enkelt, konsistent versjon av informasjonen.
2. **Automatiser testing og feildeteksjon:** Konstruksjonsdata inneholder alltid feil, enten det dreier seg om unøyaktigheter i modeller, beregningsfeil eller utdaterte spesifikasjoner. Regelmessig testing av data og eliminering av gjentakende feil kan forbedre datakvaliteten betydelig. Som en del av DataOps må du implementere automatiserte kontroller og valideringsmekanismer som overvåker at dataene er korrekte, analyserer feil og identifiserer mønstre, samt fanger opp og håndterer systemfeil i hver eneste arbeidsflyt. Jo høyere grad av automatisert validering, desto høyere blir den generelle datakvaliteten og desto lavere blir sannsynligheten for feil i sluttfasen.
3. **Data bør testes på samme måte som programkode:** De fleste bygningsapplikasjoner er basert på databehandling, men kontrollen av dem overlates ofte til sekundære roller. Hvis maskinlæringsmodeller trenes opp på unøyaktige data, fører det til feilaktige spådommer og økonomiske tap. Innen DataOps bør data underkastes den samme strenge kontrollen som programvarekode: logiske kontroller, stresstester og evaluering av modellenes oppførsel når inngangsverdiene endres. Bare validerte og pålitelige data kan brukes som grunnlag for ledelsesbeslutninger.
4. **Dataobservabilitet uten at det går på bekostning av ytelsen:** Dataovervåking er ikke bare en samling beregninger, men et strategisk kvalitetsstyringsverktøy. For at DataOps skal fungere effektivt, må observerbarhet bygges inn i alle faser av datahåndtering, fra design til drift. Samtidig er det viktig at overvåkingen ikke gjør systemet tregere. I byggeprosjekter er det avgjørende ikke bare å samle inn data, men også å gjøre det på en slik måte at arbeidet til fagfolkene (f.eks. konstruktører) som skaper dataene, ikke forstyrres på noen måte. Denne balansen gjør det mulig å kontrollere datakvaliteten uten at det går på bekostning av produktiviteten.

DataOps er ikke en ekstra byrde for dataforskerne, men selve ryggraden i arbeidet deres. Ved å implementere DataOps kan byggfirmaer gå fra kaotisk datahåndtering til et effektivt økosystem der

dataene jobber for virksomheten.

VectorOps representerer neste trinn i utviklingen av DataOps, som fokuserer på behandling, lagring og analyse av flerdimensjonale vektordata (som ble omtalt i forrige kapittel). Dette er særlig relevant på områder som digitale tvillinger, nevrale nettverksmodeller og semantiske søk, som er på vei inn i byggebransjen. VectorOps baserer seg på vektordatabaser for å lagre, indeksere og søke i flerdimensjonale representasjoner av objekter på en effektiv måte.

VectorOps er det neste steget etter DataOps, som fokuserer på behandling, analyse og bruk av vektordata i konstruksjon. I motsetning til DataOps, som fokuserer på dataflyt, konsistens og kvalitet, fokuserer VectorOps på å håndtere de flerdimensjonale objektrepresentasjonene som kreves for maskinlæring.

I motsetning til tradisjonelle tilnæringer gir VectorOps deg mulighet til å oppnå mer nøyaktige objektbeskrivelser, noe som er avgjørende for digitale tvillinger, generative designsystemer og automatisk feildeteksjon i CAD-data konvertert til vektorformat. Kombinerte implementeringen av DataOps og VectorOps danner et solid grunnlag for skalerbart, automatisert arbeid med store mengder informasjon - fra klassiske tabeller til semantisk rike romlige modeller.

Neste steg: fra kaotisk lagring til strukturert lagring

Tradisjonelle tilnæringer til datalagring fører ofte til at det opprettes ulike "informasjonssiloer" der viktig innsikt ikke er tilgjengelig for analyse og beslutningstaking. Moderne lagringskonsepter, som Data Warehouse, Data Lake og hybrider av disse, gjør det mulig å forene ulik informasjon og gjøre den tilgjengelig på en sentralisert måte for datastrømming og business intelligence. Det er ikke bare viktig å velge riktig lagringsarkitektur, men også å implementere Data Governance) og Dataminimalisme) for å forhindre at lagringsanleggene blir ukontrollerbare datasumper).

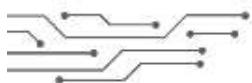
For å oppsummere denne delen er det verdt å fremheve de viktigste praktiske trinnene som vil hjelpe deg med å anvende konseptene vi har diskutert, i det daglige arbeidet:

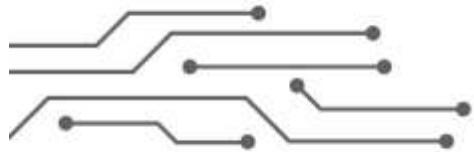
- Velg effektive formater for datalagring
 - Gå bort fra CSV og XLSX og over til mer effektive formater (Apache Parquet, ORC) for lagring av store datamengder
 - Implementere et system for versjonering av data for å spore endringer
 - Bruk metadata til å beskrive informasjonens struktur og opprinnelse
- Skap en enhetlig dataarkitektur for bedriften
 - Sammenligne ulike lagringsarkitekturen: RDBMS, DWH og Data Lake. Velg den som best oppfyller dine behov for skalerbarhet, kildeintegrasjon og analytisk behandling
 - Utform et prosesskart for uthenting, lasting og transformering av data (ETL) fra ulike kilder for oppgavene dine. Bruk visualiseringsverktøy som Miro, Lucidchart eller Draw.io til å visualisere viktige trinn og integrasjonspunkter.
- Implementere datastyringspraksis og dataminimalisme

- Følg Dataminimalisme-tilnærmingen - bare lagre og behandle det som virkelig er verdifullt
- Implementere prinsipper for datastyring - definere ansvar for data, sikre kvalitet og åpenhet
- Lær mer om retningslinjer for datahåndtering og DataOps-konsepter, VectorOps
- Definere datakvalitetskriterier og prosedyrer for datavalidering innenfor DataOps

Velorganisert datalagring skaper grunnlaget for sentralisering av bedriftens analytiske prosesser. Overgangen fra kaotisk opphopning av filer til strukturert lagring gjør det mulig å gjøre informasjon til en strategisk ressurs som bidrar til å ta velbegrunnede beslutninger og effektivisere forretningsprosessene.

Når prosessene for datainnsamling, -transformasjon, -analyse og strukturert lagring er automatisert og standardisert, er neste trinn i den digitale transformasjonen fullverdig håndtering av Big Data.





IX DEL

STORDATA, MASKINLÆRING OG PREDIKSJONER

Den niende delen fokuserer på stordata, maskinlæring og prediktiv analyse i byggebransjen. Den utforsker overgangen fra intuitiv beslutningstaking til objektive analyser basert på historiske data. Praktiske eksempler brukes til å demonstrere stordataanalyse i byggebransjen - fra analyse av datasettet for byggetillatelser i San Francisco til behandling av CAD - prosjekter med millioner av elementer. Det legges særlig vekt på maskinlæringsmetoder for å forutsi kostnader og tidsplaner for byggeprosjekter, med en detaljert diskusjon av lineær regresjon og k-nærmeste naboer-algoritmer. Det vises hvordan strukturerte data danner grunnlaget for prediktive modeller for å vurdere risiko, optimalisere ressurser og forbedre effektiviteten i prosjektstyringen. Det gis også anbefalinger for valg av representative datautvalg, og det forklares hvorfor det ikke alltid er nødvendig med store datasett for å gjennomføre effektive analyser.

KAPITTEL 9.1.

STORDATA OG ANALYSE AV DISSE

Stordata i byggebransjen: fra intuisjon til forutsigbarhet

Begrepet "big data" har ikke en streng definisjon. Konseptet dukket opprinnelig opp da informasjonsvolumet begynte å overstige mulighetene til tradisjonelle metoder for behandling. I dag har datavolumet og kompleksiteten i mange bransjer, inkludert bygg og anlegg, økt så mye at det ikke passer inn i det lokale minnet til datamaskiner og krever bruk av ny teknologi for å behandle det.

Essensen i arbeidet med Big Data er ikke bare lagring og behandling, men også prediktive evner. I byggebransjen åpner Big Data veien fra intuitive beslutninger basert på subjektiv tolkning av tabeller og visualiseringer (som diskutert tidligere) til informerte prognosenter støttet av reelle observasjoner og statistikk.

I motsetning til hva mange tror, er ikke målet med å jobbe med stordata å "få en maskin til å tenke som et menneske", men å bruke matematiske modeller og algoritmer til å analysere enorme datamengder for å identifisere mønstre, forutsi hendelser og optimalisere prosesser.

Big Data er ikke en kald verden av algoritmer uten menneskelig innflytelse. Tvert imot fungerer stordata i samspill med våre instinkter, feil og kreativitet. Det er ufullkommenheten i menneskelig tenkning som gjør det mulig for oss å finne ikke-standardiserte løsninger og gjøre gjennombrudd.

I takt med utviklingen av digital teknologi har byggebransjen begynt å ta i bruk databehandlingssteknikker som kommer fra IT-sektoren. Takket være verktøy som Pandas og Apache Parquet kan strukturerde og ustukturerte data kombineres, noe som forenkler tilgangen til informasjon og reduserer analysetapet, mens store datasett fra dokumenter eller DAK-prosjekter (fig. 9.2-10 - fig. 9.2-12) gjør det mulig å samle inn, analysere og forutsi data i alle faser av prosjektets livssyklus.

Big Data har en transformativ innvirkning på byggebransjen og kan potensielt påvirke den på en rekke ulike måter. Bruken av Big Data-teknologi gir resultater på en rekke viktige områder, blant annet på følgende områder

- **Analyse av investeringspotensialet** - prognosenter for lønnsomhet og tilbakebetalingstid for prosjekter basert på data fra tidligere anlegg.
- **Forutseende vedlikehold** - identifisering av sannsynlige feil på utstyret før de faktisk oppstår, noe som reduserer nedetiden.
- **Optimalisering av forsyningsskjeden** - forutsi forstyrrelser og forbedre logistikkens effektivitet.
- **Energieffektivitetsanalyser** - hjelp til utforming av lavenergibygg.
- **Sikkerhetsovervåking** - bruk av sensorer og bærbare enheter for å overvåke forholdene på byggeplassen.
- **Kvalitetskontroll** - sanntidsovervåking av overholdelse av prosessstandarder.

Bemanningsstyring - resultatanalyse og prognosenter bemanningsbehov.

Det er vanskelig å finne et område i bygg- og anleggsbransjen der dataanalyseprediksjoner ikke er etterspurt. Den største fordelen med prediksjonsalgoritmer er at de er selvlærende og stadig blir bedre etter hvert som dataene akkumuleres.

I nær fremtid vil kunstig intelligens ikke bare hjelpe byggherrer, men også ta viktige beslutninger - fra designprosesser til drift av bygninger.

Mer om hvordan prediksjoner genereres og læringsmodeller brukes, kommer vi tilbake til i neste del av boken, "Maskinlæring og prediksjoner".

Overgangen til fullverdig arbeid med stordata krever en endring i selve tilnærmingen til analyse. Mens de klassiske systemene vi har diskutert så langt, fokuserte på årsak-virkning-sammenhenger, skifter stordataanalysen fokus til søken etter statistiske mønstre og korrelasjoner, som kan avsløre skjulte sammenhenger og forutsi oppførselen til objekter selv uten en fullstendig forståelse av alle faktorer.

Spørsmål om gjennomførbarheten av stordata: korrelasjon, statistikk og datautvalg

Tradisjonelt sett var byggingen basert på subjektive hypoteser og personlig erfaring. Ingeniørene antok - med en viss grad av sannsynlighet - hvordan materialet ville oppføre seg, hvilke belastninger konstruksjonen ville tåle, og hvor lenge prosjektet ville vare. Disse antakelsene ble testet i praksis, ofte på bekostning av tid, ressurser og fremtidig risiko.

Med inntoget av stordata er tilnærmingen i ferd med å endre seg dramatisk: Beslutninger tas ikke lenger på grunnlag av intuitive magefølelser, men som et resultat av analyser av store datasett. Bygg og anlegg er gradvis i ferd med å gå fra å være en intuisjonskunst til å bli en presis vitenskap om prediksjon.

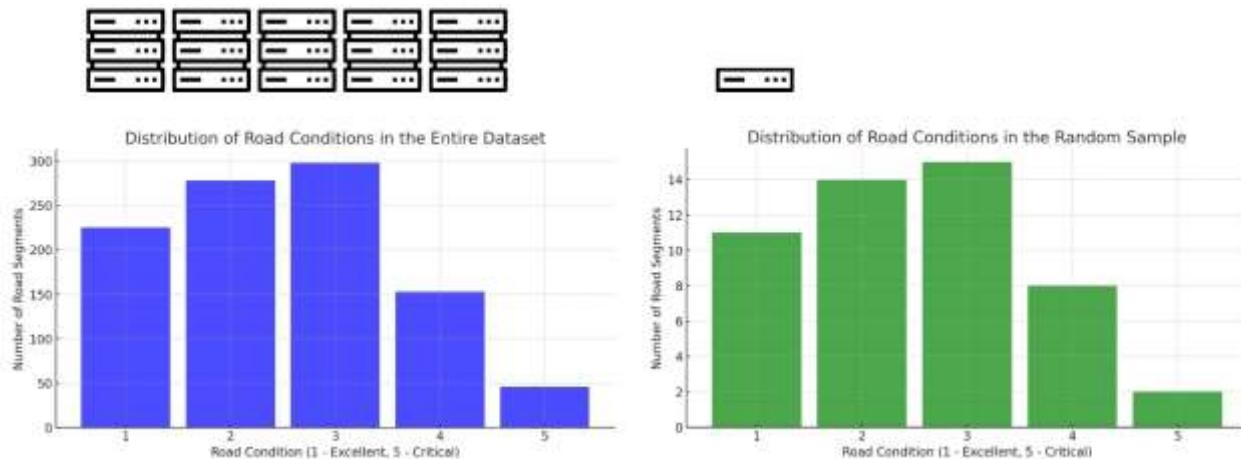
Overgangen til ideen om å bruke stordata reiser uunngåelig et viktig spørsmål: Hvor kritisk er datamengden, og hvor mye informasjon er egentlig nødvendig for pålitelig prediktiv analyse? Den utbredte oppfatningen om at "jo mer data, jo høyere nøyaktighet" viser seg ikke alltid å være statistisk gyldig i praksis.

Allerede i 1934 beviste statistikeren Jerzy Neumann [145] at nøkkelen til nøyaktigheten i statistiske sluttninger ikke så mye ligger i datamengden som i representativiteten og tilfeldigheten i utvelgelsen.

Dette gjelder spesielt i byggebransjen, der store datamengder samles inn ved hjelp av IoT -sensorer, skannere, overvåkningskameraer, droner og til og med CAD-modeller i flere formater -modeller, noe som øker risikoen for blinde flekker, ekstremverdier og dataforvrengninger.

La oss ta et eksempel med tilstandsovervåking av veidekke. Et komplett datasett med alle veistrekninger kan ta X GB og ta omtrent en dag å behandle. Samtidig vil et tilfeldig utvalg som bare omfatter hver 50. veistrekning, bare ta X/50 GB og kunne behandles på en halvtime, samtidig som det gir like nøyaktige

estimater for visse beregninger (figur 9.1-1).



Figur 9.1-1 Histogrammer over veidekkets tilstand: hele datasettet og stikkprøver viser identiske resultater.

Nøkkelen til en vellykket dataanalyse er derfor ofte ikke mengden data, men utvalgets representativitet og kvaliteten på behandlingsmetodene som brukes. Overgangen til tilfeldig utvelgelse og en mer selektiv tilnærming krever et skifte i tankegangen i byggebransjen. Historisk sett har selskaper fulgt logikken "jo mer data, jo bedre", i den tro at det å dekke alle mulige indikatorer ville maksimere nøyaktigheten.

Denne tilnærmingen minner om en populær misoppfatning fra prosjektledelse: "Jo flere spesialister jeg tiltrekker meg, desto mer effektivt blir arbeidet". Men akkurat som med menneskelige ressurser er det kvalitet og verktøy som er viktigere enn kvantitet. Hvis man ikke tar hensyn til sammenhengene (korrelasjonene) mellom data eller prosjektdeltakere, kan økt volum bare føre til støy, forvrengning, duplisering og unødvendig sløsing.

Til syvende og sist viser det seg ofte at det er mye mer produktivt å ha et mindre, men kvalitativt godt forberedt datasett som kan gi stabile og fornuftige prognosør, enn å basere seg på massiv, men kaotisk informasjon som inneholder mange motstridende signaler.

For store datamengder garanterer ikke bare ikke større nøyaktighet, men kan også føre til forvrengte konklusjoner på grunn av støy, overflødige funksjoner, skjulte korrelasjoner og irrelevant informasjon. Under slike omstendigheter øker risikoen for overtilpasning av modeller, og påliteligheten til de analytiske resultatene reduseres.

I byggebransjen er det en stor utfordring å finne den optimale mengden og kvaliteten på dataene. Når man for eksempel overvåker tilstanden til betongkonstruksjoner, kan tusenvis av sensorer og innsamling av informasjon hvert minutt overvelde lagrings- og analysesystemet. Men hvis du utfører korrelasjonsanalyse og velger ut de 10% mest informative sensorene, kan du få nesten like nøyaktige prediksjoner, samtidig som du bruker mange ganger, noen ganger titalls og hundrevis av ganger, mindre ressurser.

Ved å bruke en mindre delmengde data reduseres både lagringsbehovet og prosesseringstiden, noe som reduserer kostnadene ved lagring og analyse av data betraktelig og ofte gjør stikkprøver til en ideell løsning for prediktiv analyse, spesielt i store infrastrukturprosjekter eller når man jobber i sanntid. Til syvende og

sist er det ikke mengden innsamlede data som avgjør hvor effektive byggeprosessene er, men kvaliteten på analysen av dem. Uten en kritisk tilnærming og nøyne analyse kan data føre til feilaktige konklusjoner.

Etter en viss mengde data gir hver nye informasjonsenhet mindre og mindre nyttige resultater. I stedet for å samle inn informasjon i det uendelige, er det viktig å fokusere på representativitet og analysemetoder (figur 9.2-2).

Dette fenomenet er godt beskrevet av Allen Wallis [146], som illustrerer bruken av statistiske metoder ved hjelp av et eksempel på testing av to alternative prosjekttiltak i den amerikanske marinens tilsidesetning.

Marinen testet to alternative prosjekttilkonsruksjoner (A og B) ved å gjennomføre en rekke parvise runder. I hver runde får A en 1 eller 0, avhengig av om den presterer bedre eller dårligere enn B, og omvendt. Den statistiske standardmetoden innebærer at man gjennomfører et fast antall forsøk (f.eks. 1000) og kårer vinneren basert på en prosentfordeling (f.eks. hvis A får 1 i mer enn 53% av tilfellene, regnes den som best). Da Allen Wallis diskuterte et slikt problem med kaptein Garrett L. Schuyler i marinens tilsidesetning, innvendte kapteinen at en slik test, for å sitere Allen, kunne være ubruklig. Hadde en klok og erfaren artillerioffiser som Schuyler vært på stedet, ville han etter de første hundre [skuddene] ha sett at eksperimentet ikke trengte å bli avsluttet, enten fordi den nye metoden er klart underlegen eller fordi den er klart bedre enn det man håpet på [146].

- U.S. Government Statistical Research Group ved Columbia University, perioden under andre verdenskrig

Dette prinsippet er mye brukt i ulike bransjer. Innen medisin, for eksempel, gjennomføres kliniske studier av nye legemidler på tilfeldige utvalg av pasienter, noe som gjør det mulig å oppnå statistisk signifikante resultater uten å måtte teste legemiddelet på hele befolkningen på planeten. I økonomi og sosiologi gjennomføres representative spørreundersøkelser for å gjenspeile samfunnets mening uten at det er nødvendig å intervju alle i landet.

På samme måte som myndigheter og forskningsorganisasjoner undersøker små populasjoner for å forstå generelle samfunnstrender, kan bedrifter i byggebransjen bruke tilfeldige datautvalg til å overvåke og lage prognosenter for prosjektstyring på en effektiv måte (figur 9.1-1).

Stordata kan endre tilnærmingen til samfunnsvitenskapen, men det vil ikke erstatte statistisk sunn fornuft [147].

- Thomas Landsall-Welfair, "Forecasting the nation's current mood", Significance v. 9(4), 2012 .

Fra et ressursbesparende perspektiv er det viktig å svare på spørsmålet om det er fornuftig å bruke store ressurser på å samle inn og behandle enorme datasett når man kan bruke et mye mindre og billigere testdatasett som kan skaleres opp trinnvis, når man samler inn data for fremtidige prognoseter og

beslutninger. Effektiviteten ved tilfeldig utvelgelse viser at bedrifter kan redusere kostnadene med titusener eller til og med tusenvis av ganger på innsamling og opplæring av modeller ved å velge datainnsamlingsmetoder som ikke krever omfattende dekning, men som likevel gir tilstrekkelig nøyaktighet og representativitet. Denne tilnærmingen gjør det mulig for selv små selskaper å oppnå resultater på nivå med store selskaper ved hjelp av betydelig mindre ressurser og datamengder, noe som er viktig for selskaper som ønsker å optimalisere kostnadene og fremskynde informerte beslutninger ved hjelp av små ressurser. I de følgende kapitlene kan du se eksempler på analytics og prediktiv analyse basert på offentlige datasett ved hjelp av stordataverktøy.

Stordata: analyse av data fra San Franciscos datasett med en million byggetillatelser

Arbeid med åpne datasett gir en unik mulighet til å praktisere prinsippene som er diskutert i de foregående kapitlene: fornuftig utvelgelse av funksjoner, representativt utvalg, visualisering og kritisk analyse. I dette kapittelet skal vi se nærmere på hvordan komplekse fenomener som byggeaktivitet i en storby kan undersøkes ved hjelp av åpne data - nærmere bestemt over en million byggetillatelser i San Francisco

Offentlig tilgjengelige data om over én million byggetillatelser (figur 9.1-2) (registreringer i to datasett i CSV-format) fra "San Francisco Department of Buildings" [148] gjør at vi kan bruke den rå CSV-tabellen til å analysere ikke bare byggeaktiviteten i byen,, men også til å foreta en kritisk analyse av nyere trender og historien til San Franciscos byggebransje de siste 40 årene, fra 1980 til 2019.

Kodeeksemplene som ble brukt til å lage datasettvisualiseringene (figurene 9.1-3- figurene 9.1-8), samt visuelle grafer med kode, forklaringer og kommentarer, finner du på Kaggle-plattformen ved å søke etter "San Francisco. Construction Sector 1980-2019." [149].

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998
12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004	NaN
02/18/1992	install auto fire spks.	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992
permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zips/Zipcode	Location
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	(37.7903468760490, -122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	(37.72025851600388, -122.4644245667462)
198411	9202396	09/15/1992	9000.0	NaN	apartments	(37.79506002552974, -122.39593224461605)

Fig. 9.1-2 Datasettene inneholder informasjon om utstedte byggetillatelser med ulike objektattributter.

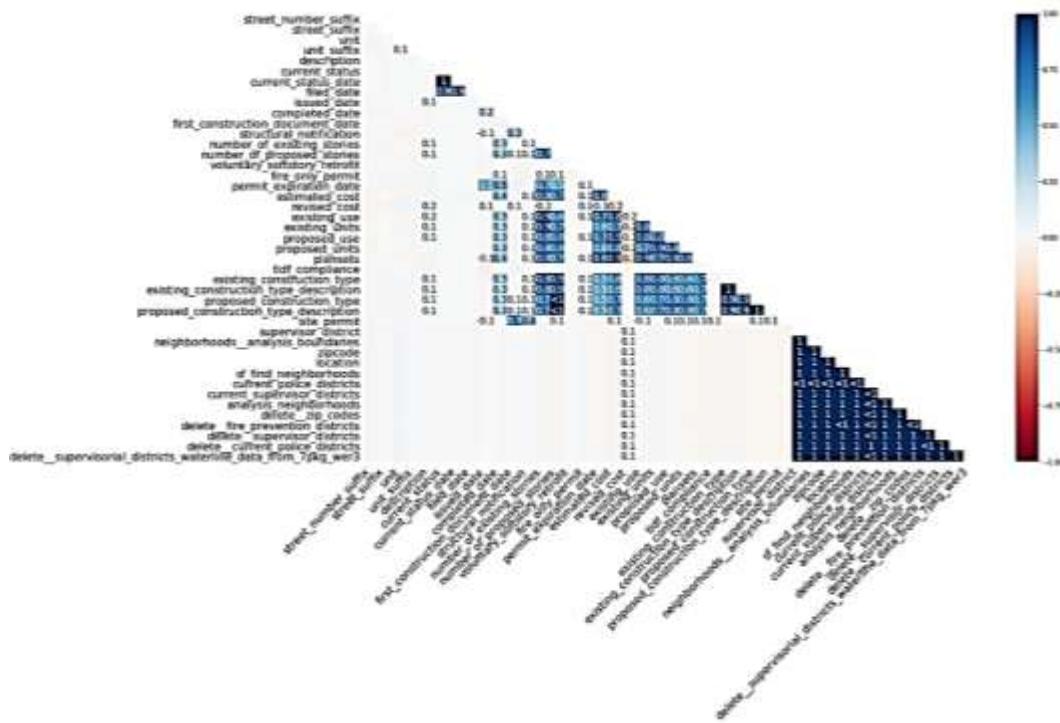


Fig. 9.1-3 Et varmekart (Pandas og Seaborn) som visualiserer alle attributtene i et datasett og hjelper til med å identifisere relasjoner mellom attributtpar.

Ingen trender eller konklusjoner fremgår av tabellen fra San Francisco Department of Buildings (figur 9.1-2). Tørre tall i tabellform er ikke noe beslutningsgrunnlag. For å gjøre dataene visuelt forståelige, slik det er beskrevet i detalj i kapitlene om datavisualisering, bør de visualiseres ved hjelp av de ulike bibliotekene som er omtalt i bokens syvende del om temaet "ETL og visualisering av resultater som grafer".

Ved å analysere data, ved hjelp av Pandas DataFrame og Python-visualiseringsbibliotekene, på verdien av 1 137 695 tillatelser [148], kan vi konkludere med at byggeaktiviteten i San Francisco er nært knyttet til økonomiske konjunkturer, spesielt i den blomstrende teknologibransjen i Silicon Valley (figur 9.1-4).

Økonomiske opp- og nedgangstider har en betydelig innvirkning på antall og verdi av byggeprosjekter. Den første toppen i byggeaktiviteten sammenfalt for eksempel med elektronikkboomen på midten av 1980-tallet (brukt Pandas og Matplotlib), og de påfølgende toppene og nedgangene var knyttet til dotcom-boblen og teknologiboomen de siste årene.

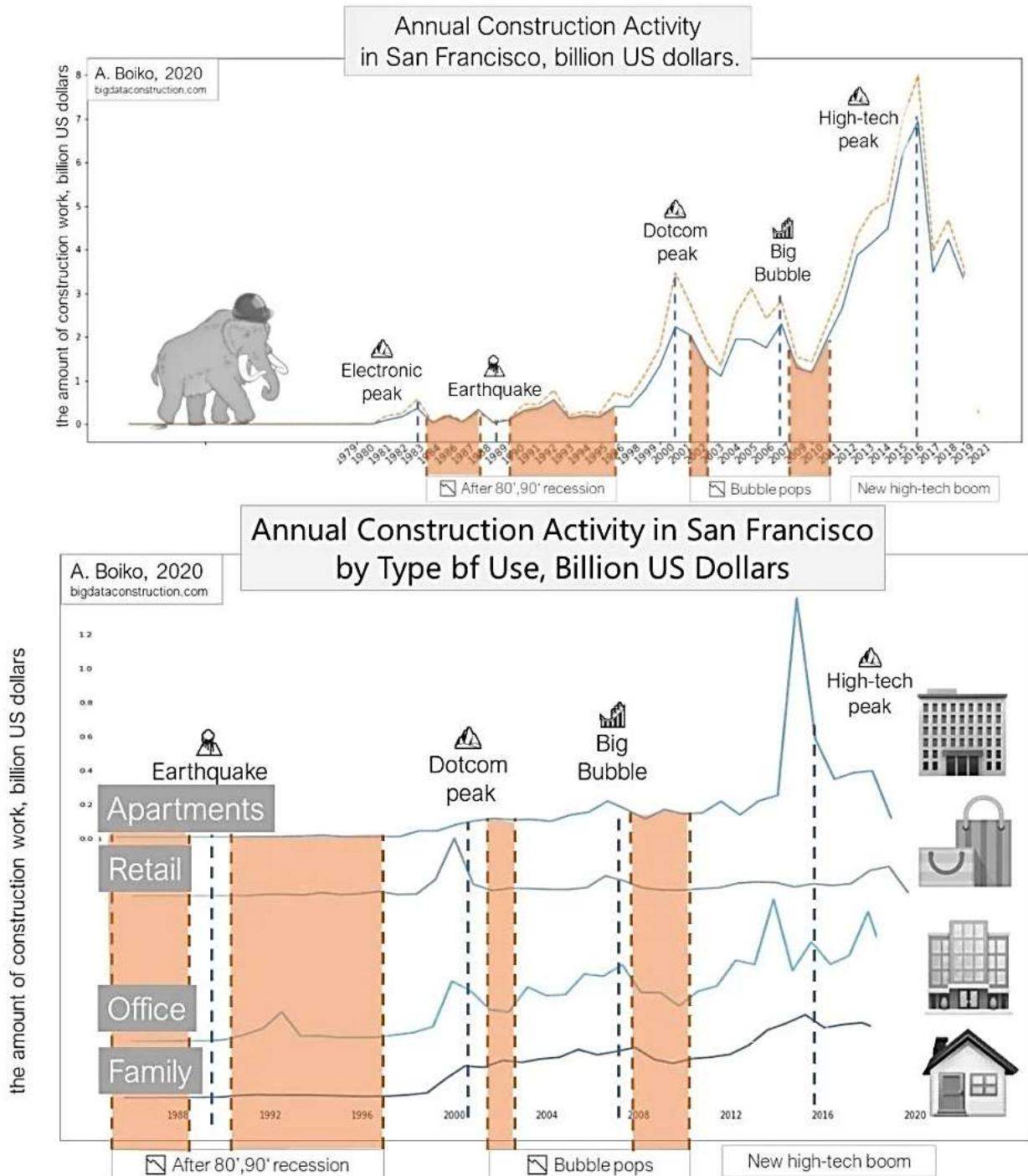
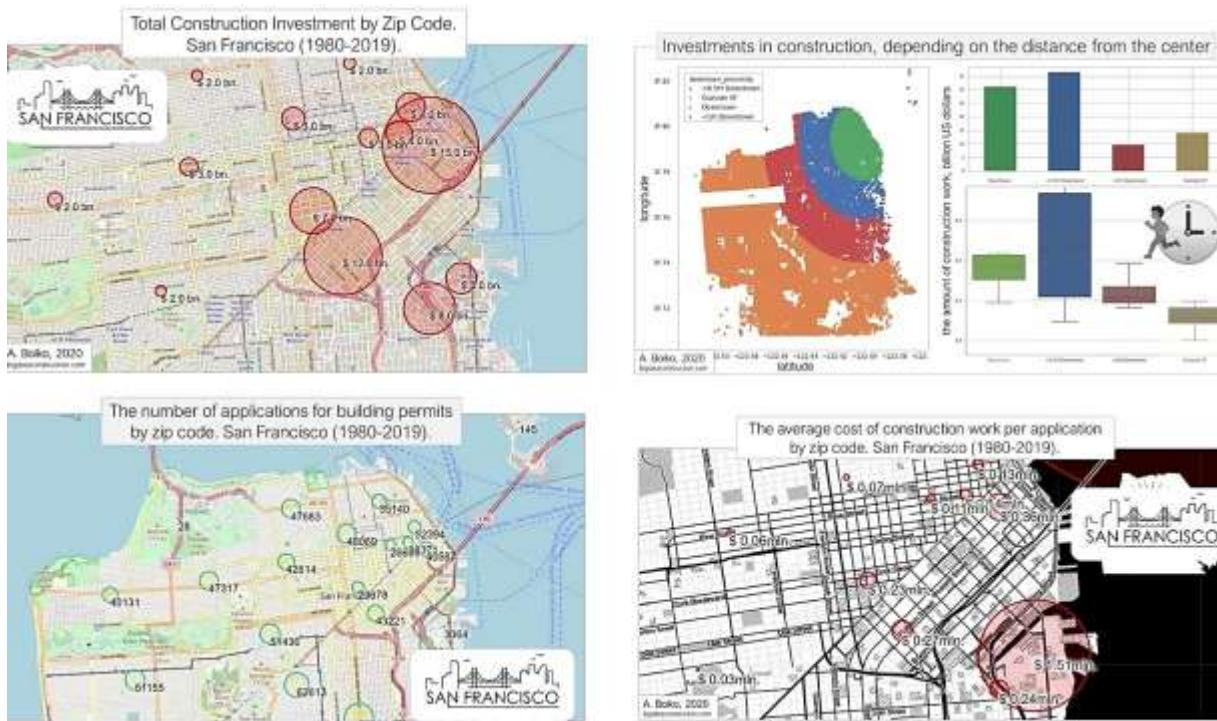


Fig. 9.1-4 I San Franciscos eiendomssektor korrelerer investeringene med den teknologiske utviklingen i Silicon Valley.

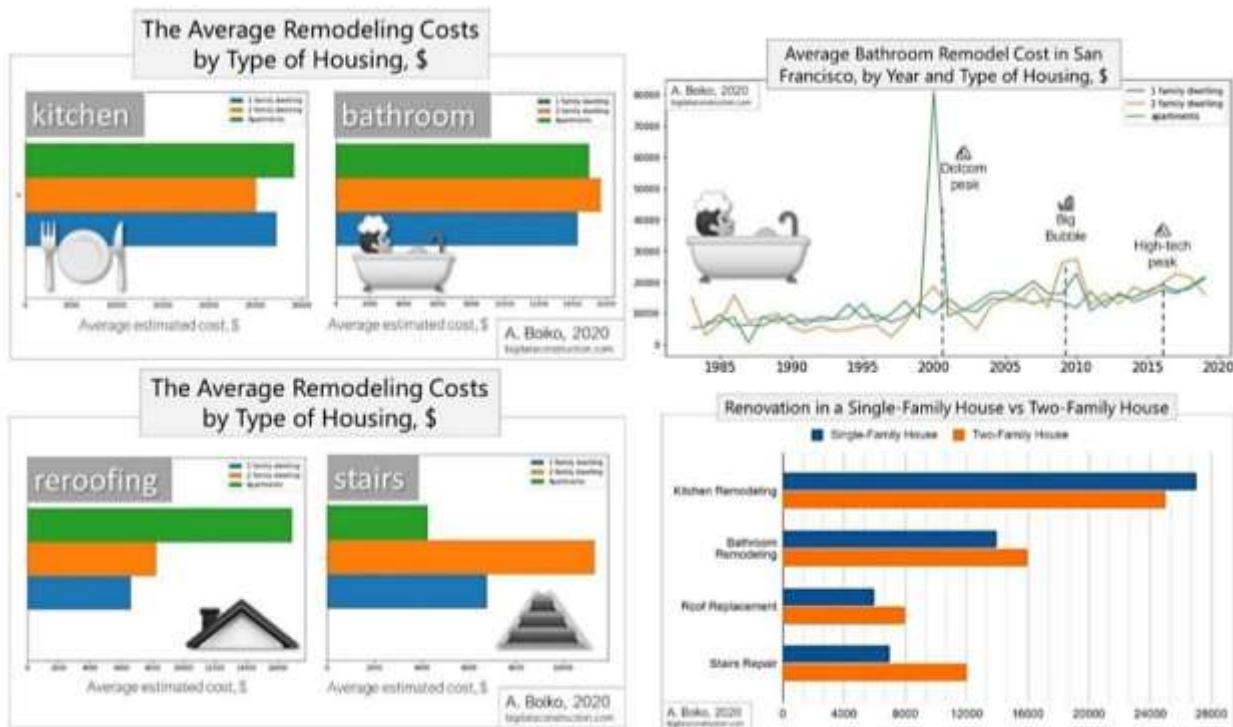
Dataanalyser tyder på at i San Francisco er mesteparten av de 91,5 milliarder dollarene som er investert i bygging og ombygging det siste tiåret - nesten 75% - konsentrert i sentrum (figur 9.1-5 - brukt Pandas og visualiseringsbiblioteket Folium) og innenfor en radius på 2 km fra sentrum, noe som gjenspeiler den høye investeringstettheten i disse sentrale områdene.

Gjennomsnittskostnaden for byggetillatelser varierer betydelig fra bydel til bydel, og søknader i sentrum koster tre ganger så mye som utenfor sentrum på grunn av høyere kostnader for tomt, arbeidskraft, materialer og strenge byggeforskrifter som krever dyrere materialer for å forbedre energieffektiviteten.



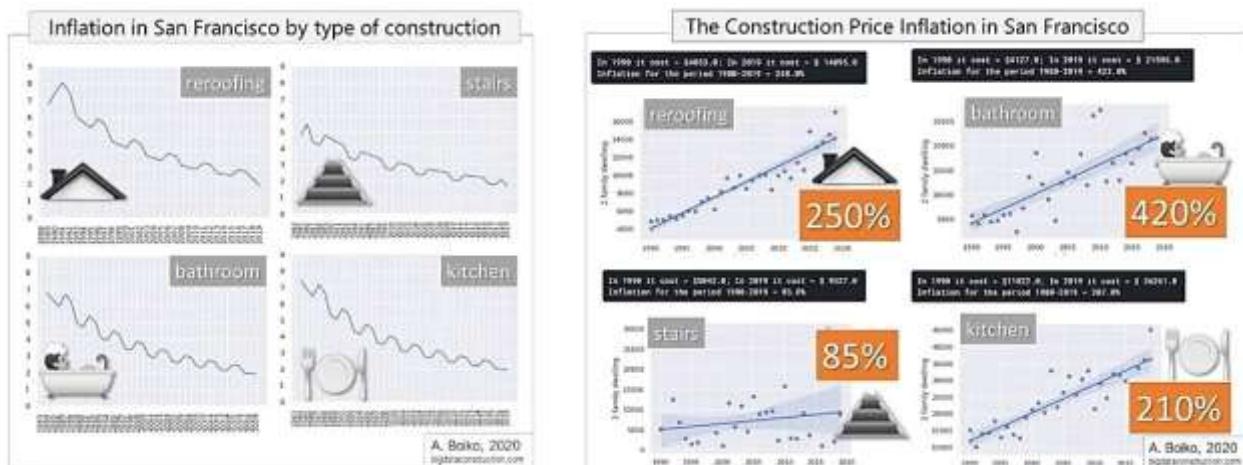
Figur 9.1-5 I San Francisco er 75 prosent av byggeinvesteringene (91,5 milliarder dollar) konsentrert i sentrum.

Datasettet gjør det også mulig å beregne gjennomsnittlige reparasjonspriser ikke bare etter hustype, men også etter bydeler og individuelle adresser (postnumre). I San Francisco viser dynamikken i oppussingskostnadene tydelige trender for ulike typer oppussing og boliger (figur 9.1-6 brukt Pandas og Matplotlib). Kjøkkenoppussing er merkbart dyrere enn baderomsoppussing: Den gjennomsnittlige kjøkkenoppussingen i en enebolig koster rundt 28 000 dollar, sammenlignet med 25 000 dollar i en tomannsbolig.



Figur 9.1-6 I SF koster oppussing av kjøkken nesten dobbelt så mye som oppussing av bad, og huseiere må sette av \$ 350 hver måned i 15 år for å dekke kostnadene ved større reparasjoner.

Byggekostnadsveksten i San Francisco gjennom årene kan spores ved å analysere data gruppert etter boligtype og år (figur 9.1-7 - brukt av Pandas og Seaborn), som viser en jevn økning i de gjennomsnittlige reparasjonskostnadene siden 1990 og avslører kortsiktige treårssykluser i kostnadene for reparasjoner av flerfamiliehus.



Figur 9.1-7 Fra 1980 til 2019 har kostnadene for baderomsoppussing i SF femdoblet seg, mens tak- og kjøkkenoppussing har tredoblet seg i pris og trappeoppussing kun har økt med 85%.

En studie av offentlige data fra San Francisco Building Department (figur 9.1-3) viser at byggekostnadene i byen er ekstremt variable og ofte uforutsigbare, og at de påvirkes av en rekke faktorer. Disse faktorene

inkluderer økonomisk vekst, teknologisk innovasjon og de unike kravene som ulike boligtyper stiller.

Tidligere krevde slike analyser inngående kunnskap om programmering og analyse. Men med LLM-verktøyene har prosessen blitt tilgjengelig og forståelig for et bredt spekter av fagfolk i byggebransjen, fra ingeniører i prosjekteringsavdelingene til toppledelsen.

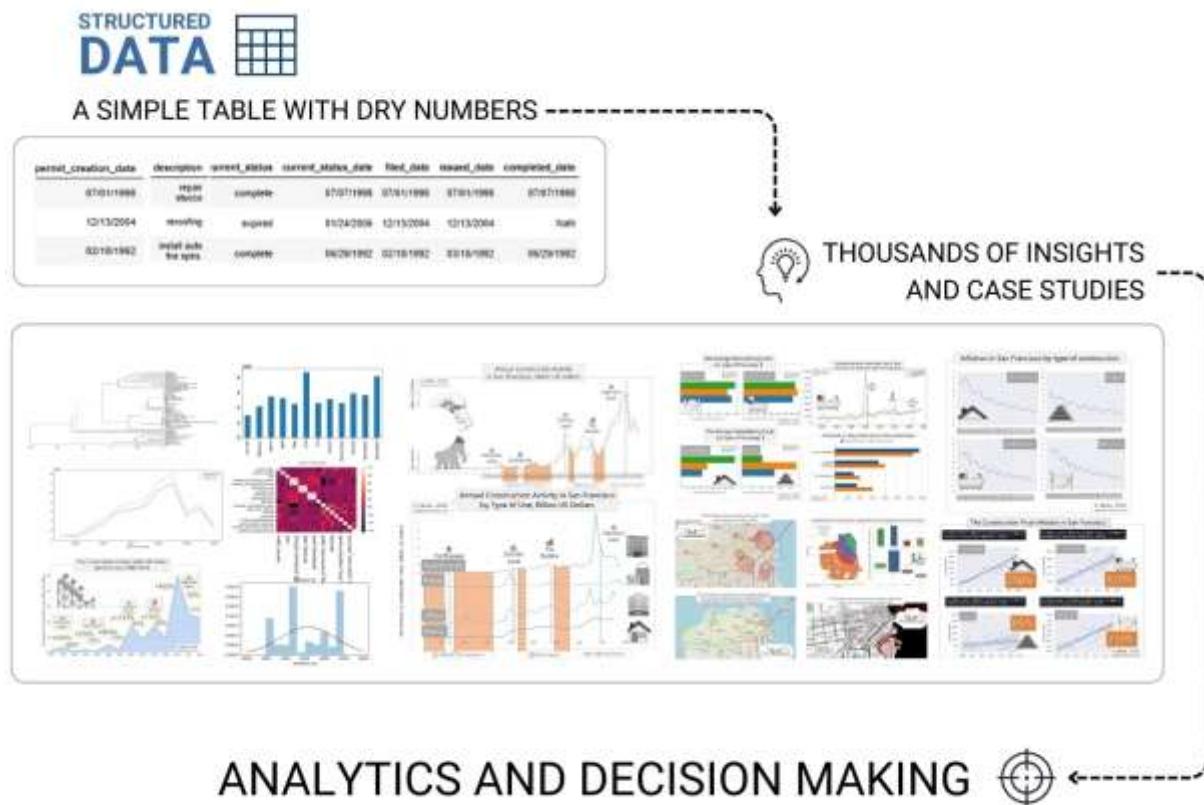


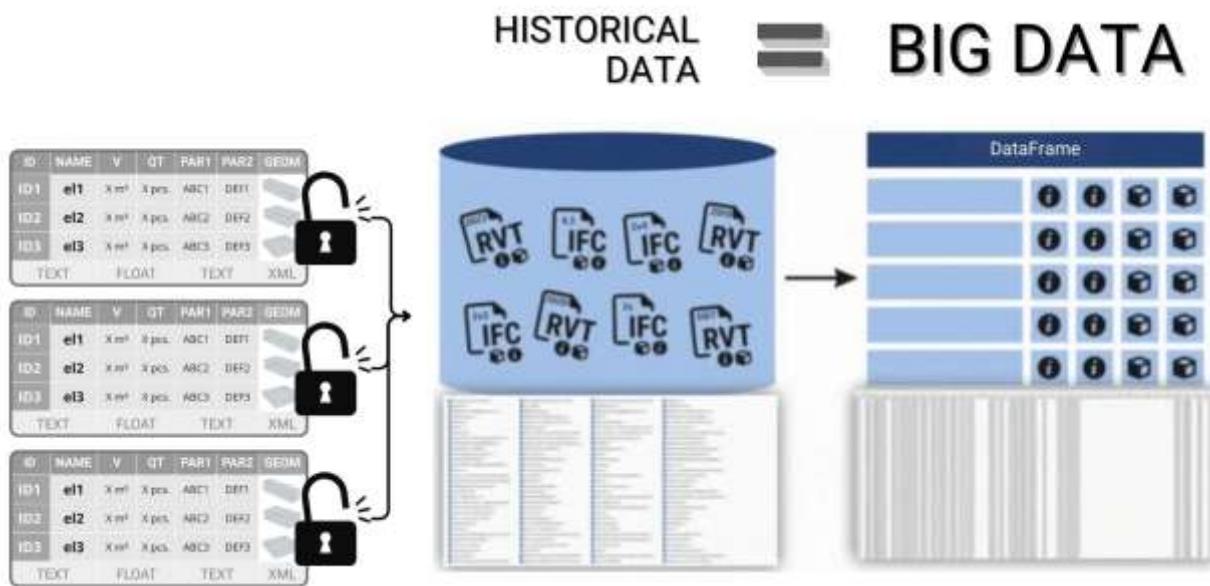
Fig. 9.1-8 Overgangen til visuelt forståelige data gjør det mulig å ta automatiserte beslutninger ved å gjenkjenne skjulte mønstre.

På samme måte som vi analyserte data fra det tabellariske datasettet "San Francisco Building Authority", kan vi visualisere og analysere et hvilket som helst datasett - fra bilder og dokumenter til IoT-data, eller data fra avleddede CAD-databaser.

Eksempel på stordata basert på CAD-data (BIM)

I det følgende eksemplet analyserer vi et stort datasett med data fra ulike DAK-verktøy (BIM). For å samle inn og opprette det store datasettet ble det brukt en spesialisert, automatisert webcrawler (skript) som var konfigurert til automatisk å søke etter og samle inn designfiler fra nettsteder som tilbyr gratis arkitekturmodeller i formatene RVT og IFC. I løpet av noen få dager fant og lastet crawleren ned 4596 IFC-filer, 6471 RVT-filer og 156 024 DWG-filer[149].

Etter å ha samlet inn prosjekter i RVT og IFC i ulike versjoner og konvertert dem til et strukturert CSV-format ved hjelp av gratis reverse engineering SDK, ble nesten 10 000 RVT- og IFC-prosjekter samlet i en stor Apache Parquet-tabellfil og lastet opp til Pandas DataFrame for analyse (figur 9.1-9).



Figur 9.1-9 Structured Data prosjektdata gjør det mulig å kombinere et hvilket som helst antall prosjekter i én enkelt todimensjonal tabell.

Dataene fra denne store samlingen inneholder følgende informasjon: IFC-filsettet inneholder ca. 4 millioner entiteter (rader) og 24 962 attributter (kolonner), og RVT-filsettet, som består av ca. 6 millioner entiteter (rader), inneholder 27 025 ulike attributter (kolonner).

Disse informasjonssettene (figur 9.1-10) dekker millioner av elementer, og for hvert av dem ble koordinatene til Bounding Box -geometrien (et rektangel som definerer grensene for et objekt i prosjektet) innhentet og lagt til i en felles tabell - koordinatene til Bounding Box -geometrien (et rektangel som definerer grensene for et objekt i prosjektet) og bilder av hvert element i PNG-format og geometrien i det åpne XML-formatet - DAE (Collada) ble opprettet.

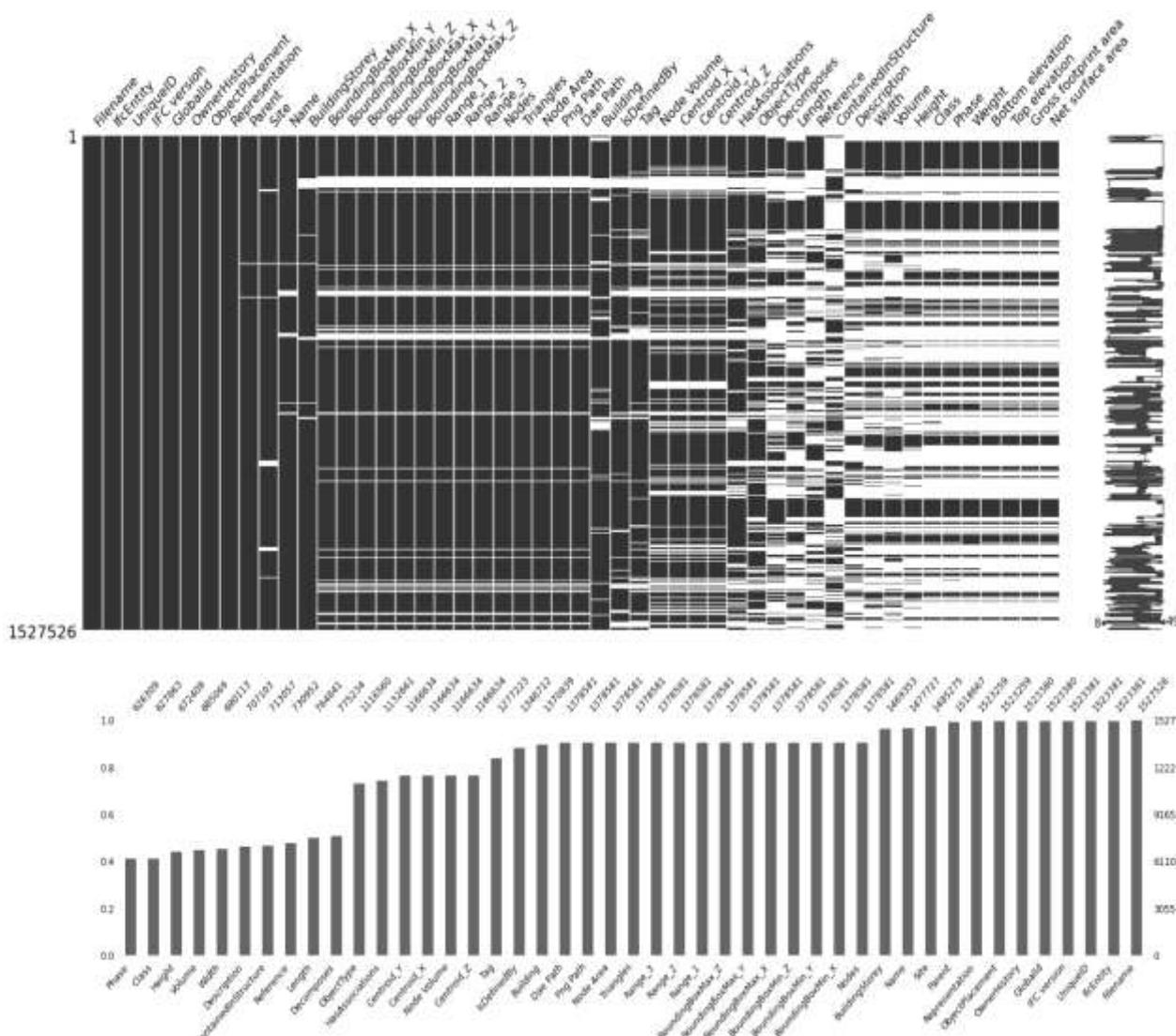


Fig. 9.1-10 Ssubset med 1,5 millioner elementer og visualisering (missingno-biblioteket) av belegget av de første 100 attributtene som et histogram.

Dermed fikk vi all informasjon om flere titalls millioner elementer fra 4596 IFC-prosjekter og 6471 RVT-prosjekter, der alle attributt-egenskapene til alle enhetselementer og deres geometri (Bounding Box) ble oversatt til en strukturert form av en enkelt tabell (DataFrame) (fig. 9.1-10 - data om dataframepopulasjoner vises som histogrammer).

Histogrammene (fig. 9.1-10, fig. 9.2-6, fig. 9.2-7) som plottes under analyseprosessen, gir en rask vurdering av datatettheten og hyppigheten av forekomsten av verdier i kolonnene. Dette gir et første innblikk i fordelingen av funksjoner, forekomsten av ekstremverdier og den potensielle nyten av individuelle attributter ved analyse og bygging av maskinlæringsmodeller.

Et eksempel på praktisk bruk av dette datasettet (fig. 9.1-10) er prosjektet "5000 IFC and RVT ". [149], som er tilgjengelig på Kaggle-plattformen. Det presenterer Jupyter Notebook med en komplett pipeline-løsning: fra forbehandling og analyse av data til visualisering av resultater ved hjelp av Python-biblioteker - pandas,

matplotlib, seaborn, folium og andre (fig. 9.1-11).



Figur 9.1-11 Eksempler på analyse av data fra CAD-formater (BIM) ved hjelp av Python-visualiseringsbiblioteker og pandas-biblioteket.

Basert på metainformasjon er det mulig å finne ut i hvilke byer bestemte prosjekter har blitt utviklet, og å vise dette på et kart (f.eks. ved hjelp av Folium-biblioteket). I tillegg gjør tidsstemplene i dataene det mulig å utforske mønstre i når filer ble lagret eller redigert: etter ukedag, tidspunkt på døgnet og måned.

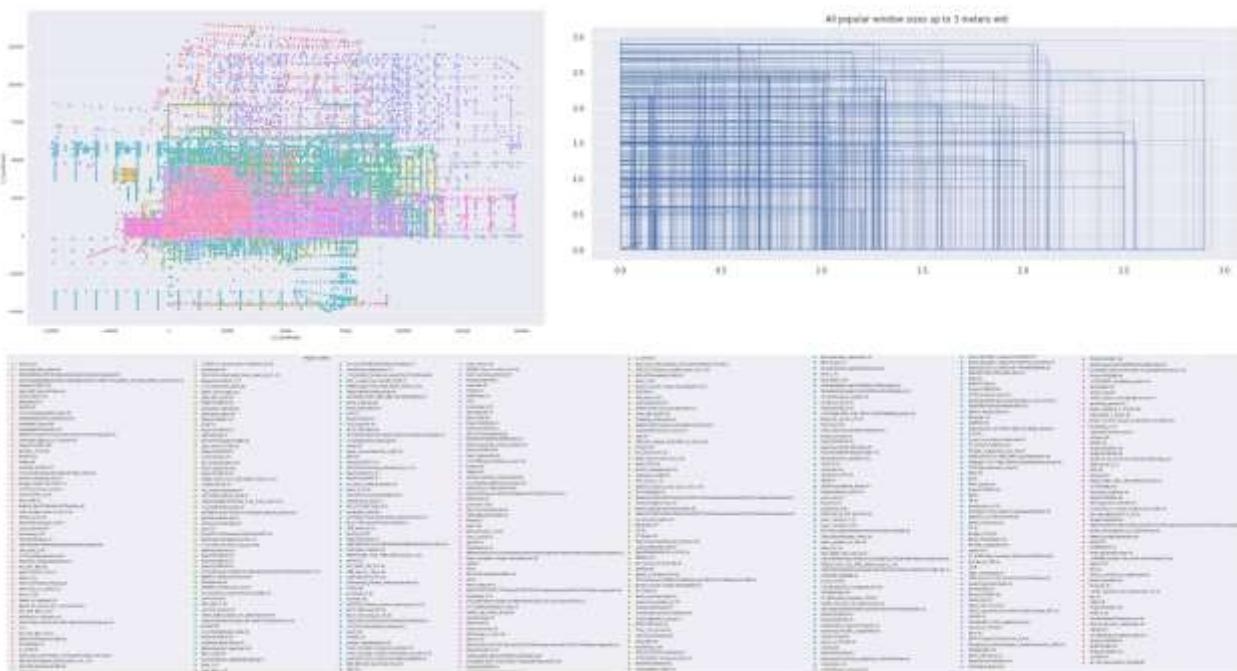


Fig. 9.1-12 Visualisering av den geometriske posisjonen til alle søyler og dimensjonene til alle vinduer opp til 3 meter i prosjekter fra listen nederst i diagrammet.

Geometriske parametere i form av Bounding Box hentet fra modellene egner seg også til aggregerte analyser. Figur 9.1-12 viser for eksempel to grafer: Den venstre viser fordelingen av avstander mellom søyler for alle prosjekter i forhold til nullpunktet, og den høyre viser dimensjonene til alle vinduer som er opptil 3

meter høye i et utvalg på titusenvis av vinduselementer (etter gruppering av hele datasettet etter parameteren "Category" med verdien "OST_Windows", "IfcWindows").

Pipeline-analysekoden for dette eksempelet og selve datasettet er tilgjengelig på Kaggle-nettstedet under tittelen "5000 IFC and RVT | DataDrivenConstruction.io projects" [149]. Denne ferdige Pipeline sammen med datasettet kan kopieres og kjøres gratis online på Kaggle eller offline i en av de populære IDE-ene: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse med PyDev-plugin, Thonny, Wing IDE, IntelliJ IDEA med Python-plugin, JupyterLab eller de populære nettverktøyene Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Den analytiske innsikten man får ved å behandle og studere enorme mengder strukturerte data, vil spille en avgjørende rolle i beslutningsprosessene i byggebransjen.

Med denne typen informasjonsanalyse basert på tidligere prosjekter kan spesialistene effektivt forutse for eksempel material- og arbeidskraftbehov og optimalisere designløsninger før byggingen starter

Men mens prosjekteringsdata eller byggetillatelser er relativt statisk informasjon som endrer seg relativt sakte, er selve byggeprosessen i ferd med å bli mettet med en rekke sensorer og IoT -enheter: kameraer, automatiserte overvåkingssystemer som overfører data i sanntid - alt dette gjør byggeplassen til et dynamisk digitalt miljø der data må analyseres i sanntid.

IoT Tingenes internett og smartkontrakter

IoT Tingenes internett representerer en ny bølge av digital transformasjon der hver enhet får sin egen IP-adresse og blir en del av et globalt nettverk. IoT er et konsept som innebærer at fysiske gjenstander kobles til Internett for å samle inn, behandle og overføre data. I byggebransjen betyr dette muligheten til å kontrollere byggeprosesser i sanntid, minimere materialavfall, forutsi slitasje på utstyr og automatisere beslutningsprosesser.

Ifølge CFMA-artikkelen "Preparing for the Future with Connected Construction" [150] vil byggebransjen gjennomgå en stor digital transformasjon i løpet av det neste tiåret, noe som vil kulminere i konseptet Connected Construction - en fullt integrert og automatisert byggeplass.

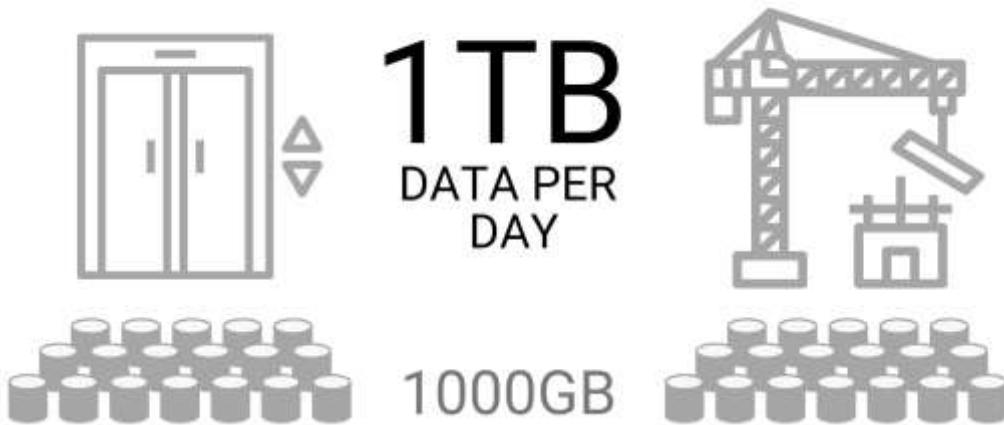


Fig. 9.1-13 IoT- eller byggeplassdataenheter kan produsere og overføre terabytes med data per dag.

En digital byggeplass innebærer at alle elementer av byggingen - fra planlegging og logistikk til arbeidsutførelse og kvalitetskontroll på byggeplassen ved hjelp av faste kameraer og quadrokoptere - integreres i ett enkelt dynamisk digitalt økosystem. Tidligere, i del 7 av denne boken, har vi allerede sett på mulighetene som ligger i Apache NiFi (figur 7.4-5), et gratis verktøy med åpen kildekode som muliggjør datastrømming i sanntid - fra innsamling fra ulike kilder til overføring til lagrings- eller analyseplattformer.

Data om fremdrift, materialforbruk, utstyrssstatus og sikkerhet overføres i sanntid til analysesystemer (fig. 9.1-13). Dette gjør det mulig å forutse potensielle risikoer, reagere raskt på avvik og optimalisere prosessene på byggeplassen. Viktige komponenter i en digital byggeplass er blant annet

- IoT -sensorer - sporing av miljøparametere, overvåking av anleggsutstyr og kontroll av arbeidsforhold.
- Digitale tvillinger - virtuelle modeller av bygninger og infrastruktur for å forutsi mulige avvik og forebygge feil.
- Automatiserte logistikksystemer - styring av forsyningsskjeden i sanntid for å redusere nedetid og kostnader.
- Robotiserte konstruksjonssystemer - bruk av autonome maskiner til å utføre rutinemessige og farlige oppgaver.

Robotisering bruken av IoT og det digitale byggeplasskonseptet Connected Site (Construction) vil ikke bare øke effektiviteten og redusere kostnadene, men også innlede en ny æra med sikkerhet, bærekraftig bygging og prediktiv prosjektstyring.

RFID (Radio Frequency Identification) er også en av de viktigste komponentene i IoT. De brukes til å identifisere og spore materialer, maskiner og til og med personell på en byggeplass, noe som gir økt åpenhet og kontroll over prosjekttressursene.

RFID-teknologi brukes til automatisk gjenkjenning av gjenstander ved hjelp av radiosignaler. Den består av tre nøkkelelementer:

- RFID - merkelapper (passive eller aktive) - inneholder en unik identifikator og festes på materialer, verktøy eller maskiner.

- Skannere er enheter som leser informasjon fra brikker og overfører den til systemet.
- Sentralisert database - lagrer informasjon om objektenes plassering, status og bevegelse.

Bruk av RFID i bygg- og anleggsbransjen:

- Automatisk materialregnskap - merkelapper på ferdigbetongprodukter, armering eller sandwichelementpakker gjør det mulig å kontrollere lagerbeholdningen og forhindre tyveri.
- Kontroll av personalets arbeidstid - RFID - personalbrikkene registrerer start- og sluttidspunkt for skiftet og gir oversikt over arbeidstiden.
- Utstyrsovervåking - RFID - systemet sporer utstyrets bevegelser, forhindrer driftsstans og forbedrer logistikkens effektivitet.

Som et supplement til denne teknologipakken finnes det blokkjedebaserte smartkontrakter som automatiserer betalinger, leveransekontroll og kontraktsoverholdelse uten behov for mellommenn, noe som reduserer risikoen for svindel og forsinkelser.

I dag er smartkontrakter, i mangel av en felles datamodell, rett og slett kode som deltakerne blir enige om. Med en datasentrisk tilnærming er det imidlertid mulig å lage en felles modell for kontraktsparametere, kode den i en blokkjede og automatisere oppfyllelsen av vilkårene.

I et system for styring av forsyningsskjeden kan en smartkontrakt for eksempel spore leveransen av en forsendelse ved hjelp av IoT-sensorer og RFID-tags og automatisk overføre betalingen når den ankommer. På samme måte kan en smartkontrakt på en byggeplass registrere ferdigstillelsen av en arbeidsfase - for eksempel montering av armeringsjern eller støping av et fundament - basert på data fra droner eller byggesensorer, og automatisk iverksette neste betaling til entreprenøren uten behov for manuelle kontroller og papirstifikater.

Men til tross for ny teknologi og innsatsen til internasjonale standardiseringsorganisasjoner, er det en mengde konkurrerende standarder som kompliserer IoT-landskapet.

Ifølge en Cisco-studie publisert i 2017 [151] stopper nesten 60% av IoT-initiativene (IoT) på proof-of-concept-stadiet, og bare 26% av selskapene anser IoT-prosjektene sine som fullt ut vellykkede. En tredjedel av de fullførte prosjektene oppnår dessuten ikke de oppsatte målene og blir ikke anerkjent som vellykkede selv etter implementeringen.

En av hovedårsakene er mangelen på interoperabilitet mellom plattformer som behandler data fra ulike sensorer. Resultatet er at dataene forblir isolert i separate løsninger. Et alternativ til denne tilnærmingen, som i andre lignende tilfeller (som vi har dekket i denne boken), er en arkitektur som er bygget opp rundt selve dataene som den primære ressursen.

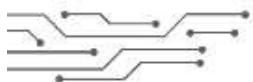
IoT-sensorer spiller en nøkkelrolle ikke bare når det gjelder å overvåke utstyrets tekniske tilstand, men også når det gjelder prediktiv analyse for å redusere risikoen på byggeplassen og forbedre den generelle prosessytelsen ved å forutsi feil og avvik.

Dataene som samles inn av IoT-sensorer og RFID-brikker, kan behandles i sanntid av maskinlæringsalgoritmer som kan oppdage uregelmessigheter og varsle ingeniører om potensielle feil på forhånd. Det kan dreie seg om alt fra mikrosprekker i betongkonstruksjoner til ukarakteristiske pauser i

driften av tårnkrarer, noe som kan indikere tekniske feil eller brudd på regelverket. Videre kan avanserte atferdsanalysealgoritmer fange opp atferdsmønstre som for eksempel kan tyde på fysisk utmattelse hos personalet, noe som bidrar til proaktiv styring av sikkerheten og de ansattes trivsel på byggeplassen.

I byggebransjen skjer ulykker og feil - enten det gjelder maskiner eller mennesker - sjeldent plutselig. De innledes vanligvis av mindre avvik som ikke blir lagt merke til. Forutseende analyser og maskinlæring gjør det mulig å oppdage disse signalene på et tidlig stadium, til og med før kritiske konsekvenser oppstår.

Mens dokumenter, prosjektfiler og data fra IoT-enheter og RFID-brikker utgjør byggeprosjektenes digitale fotavtrykk, kan maskinlæring bidra til å trekke ut nyttig innsikt fra disse. Med veksten i datamengden og demokratiseringen av datatilgangen får byggebransjen nye muligheter innen analyse, prediktiv analyse og kunstig intelligens.



KAPITTEL 9.2.

MASKINLÆRING OG PREDIKSJONER

Maskinlæring og kunstig intelligens vil endre måten vi bygger på

Databasene til de ulike systemene i byggebransjen - med sin uunngåelig forfallende og stadig mer komplekse infrastruktur - er i ferd med å bli en grobunn for fremtidige løsninger. Bedriftens servere er som en skog, rik på en biomasse av viktig informasjon, ofte gjemt under jorden, i mapper og servere. Datamassene fra de ulike systemene som skapes i dag etter bruk, etter å ha falt til bunns på serveren og etter mange års fossilisering - vil i fremtiden gi næring til maskinlæring og språkmodeller. Interne chatterom i bedriften (f.eks. en egen instans av lokalt konfigurerte ChatGPT, LLaMa, Mistral, DeepSeek) vil bygge på disse interne modellene ved hjelp av sentralisert lagring for raskt og enkelt å hente ut informasjon og generere de nødvendige grafene, instrumentbrettene og dokumentene.

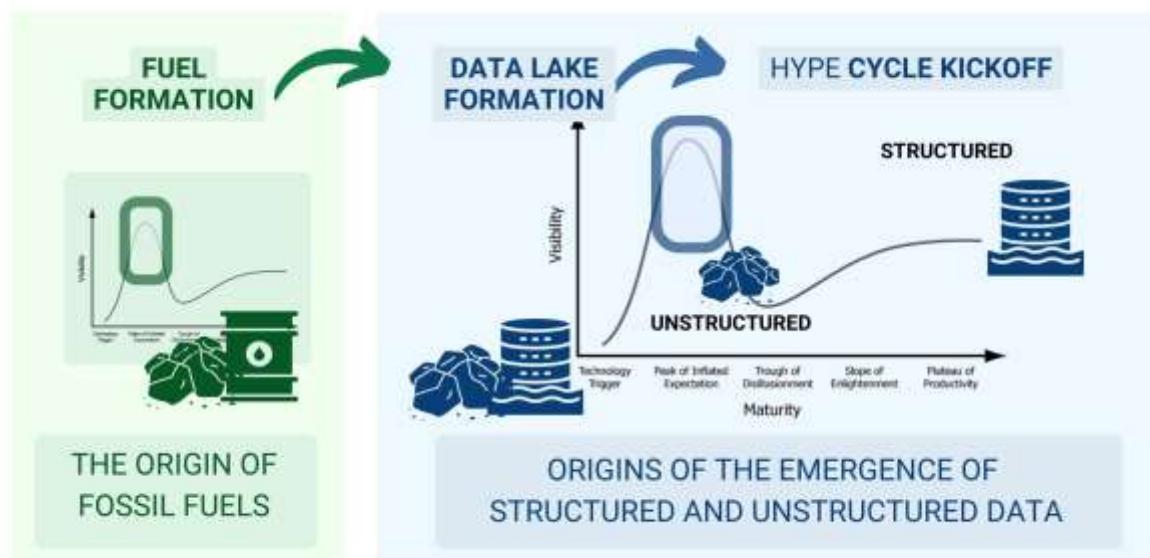


Fig. 9.2-1 Akkurat som trær blir til kull, blir også informasjon til verdifull forretningsenergi over tid under tidspress og analyse.

Fossilisering av plantemasse i kombinasjon med trykk og temperatur skaper en homogen og unikt strukturert homogen masse av trær av ulike arter som levde til ulike tider - trekull [152]. På samme måte vil informasjon som er lagret på harddisker i ulike formater og til ulike tider under trykket fra analyseavdelingene og temperaturen fra kvalitetsstyringen, til slutt danne en homogen strukturert masse av verdifull informasjon (fig. 9.2-1).

Disse lagene (eller oftere isolerte nuggets) av informasjon skapes gjennom møysommelig dataregistrering av erfarte analytikere som gradvis begynner å trekke ut verdifull informasjon fra tilsvarende irrelevant data.

I det øyeblikket disse modne datalagene ikke lenger bare "brennes" i rapporter, men begynner å sirkulere i forretningsprosessene, beriker beslutninger og forbedrer prosesser, er bedriften klar for neste skritt - overgangen til maskinlæring og kunstig intelligens (fig. 9.2-2).

Maskinlæring (ML - Machine learning) er en klasse av metoder for å løse problemer innen kunstig intelligens. Maskinlæringsalgoritmer gjenkjenner mønstre i store datasett og bruker dem til å lære seg selv. Hvert nye datasett gjør det mulig for de matematiske algoritmene å forbedre og tilpasse seg i henhold til den innhente informasjonen, noe som gjør det mulig å stadig forbedre nøyaktigheten av anbefalinger og spådommer.

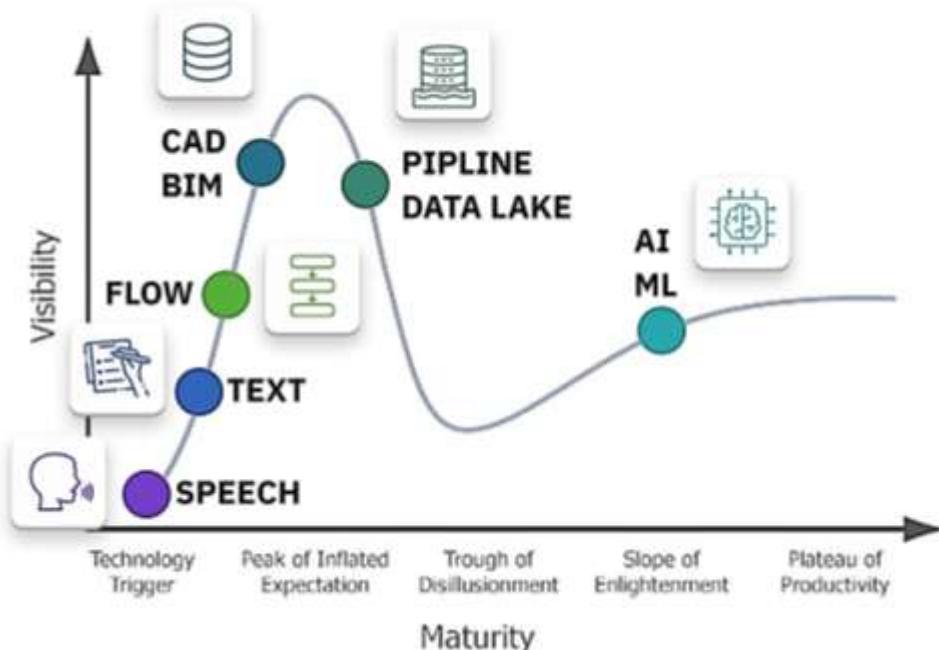


Fig. 9.2-2 Utfasingen av dataskapingsteknologier og anvendelsen av analyseverktøy åpner døren til temaet maskinlæring.

Som den innflytelsesrike administrerende direktøren for verdens største investeringsfond (som eier viktige eierandeler i nesten alle de største byggprogramvareselskapene, i tillegg til selskapene som eier den største mengden eiendom i verden [55]) sa i et intervju i 2023 - maskinlæring vil forandre byggebransjen.

AI har et enormt potensial. Det vil endre måten vi jobber på, måten vi lever på. KI og robotteknologi vil endre måten vi jobber på og måten vi bygger på, og vi vil kunne bruke KI og robotteknologi som et middel til å skape mye større produktivitet [153].

- Administrerende direktør i verdens største investeringsfond, intervju, september 2023.

Maskinlæring (ML) fungerer ved å behandle store datamengder ved hjelp av statistiske teknikker for å etterligne aspekter ved menneskelig tenkning. De fleste bedrifter har imidlertid ikke slike datasett, og hvis de har det, er de ofte ikke tilstrekkelig merket. Det er her semantiske teknologier og transfer learning, en teknikk som gjør ML mer effektiv når det gjelder små datamengder, og som vi har diskutert muligheten for i tidligere kapitler i denne delen, kan være til hjelp.

Essensen i transfer learning er at i stedet for å lære hver oppgave fra bunnen av, kan du bruke kunnskap fra

beslektede områder. Det er nødvendig å innse at mønstre og oppdagelser fra andre bransjer kan tilpasses og brukes i byggebransjen. For eksempel kan metoder for optimalisering av logistikkprosesser som er utviklet i detaljhandelen, bidra til å forbedre effektiviteten i leverandørkjedestyringen i byggebransjen. Stordataanalyse, som brukes aktivt i finansverdenen, kan brukes til kostnadsprognoser og risikostyring i byggeprosjekter. Og datasyn- og robotteknologi som utvikles i industrien, finner allerede anvendelse i automatisert kvalitetskontroll, sikkerhetsovervåking og anleggsstyring på byggeplasser.

Overføringslæring gjør det ikke bare mulig å fremskynde innføringen av innovasjoner, men også å redusere kostnadene ved utviklingen av dem ved å bruke allerede akkumulert erfaring fra andre bransjer.

$$\text{labor productivity in construction} = f(\text{AI})$$

Fig. 9.2-3 Kunstig intelligens teknologier og robotteknologi vil være fremtidens viktigste drivkraft for å øke produktiviteten i byggebransjen.

Menneskelig tenkning er organisert etter et lignende prinsipp: Vi bygger på tidligere tilegnet kunnskap for å løse nye problemer (fig. 4.4-19, fig. 4.4-20, fig. 4.4-21). Denne tilnærmingen fungerer også innen maskinlæring - ved å forenkle datamodellen og gjøre den mer elegant kan vi redusere kompleksiteten i problemet for ML-algoritmene. Dette reduserer i sin tur behovet for store datamengder og reduserer beregningskostnadene.

Fra subjektiv vurdering til statistisk prognose

Tiden da strategiske beslutninger var avhengig av den enkelte leders intuisjon (figur 9.2-4), hører fortiden til. I et stadig mer konkurransepreget og utfordrende økonomisk miljø blir en subjektiv tilnærming for risikabel og ineffektiv. Bedrifter som fortsetter å basere seg på personlige meninger i stedet for å analysere data objektivt, mister evnen til å reagere raskt på endringer.

Konkurransemiljøet krever nøyaktighet og repeterbarhet basert på data, statistiske mønstre og beregbare sannsynligheter. Beslutninger kan ikke lenger baseres på følelser, de må baseres på korrelasjoner, trender og prediktive modeller utledet fra analyser og maskinlæring. Dette er ikke bare en endring i verktøy - det er en endring i tenkningens logikk: fra antakelser til bevis, fra subjektive sannsynligheter til statistisk beregnede avvik, fra følelser til fakta.



Fig. 9.2-4 Tiden da beslutninger ble tatt av HiPPO (den best betalte medarbeiderens mening) vil bli en saga blott med inntoget av stordata og maskinlæring.

Ledere som tidligere stolte utelukkende på sine egne følelser, vil uunngåelig møte en ny virkelighet: Det er ikke lenger autoriteten som bestemmer valgene. I sentrum for ledelse står nå systemer som analyserer millioner av parametere og vektorer, identifiserer skjulte mønstre og foreslår optimale strategier.

Hovedgrunnen til at bedrifter i dag fortsatt unngår å implementere ML, er mangelen på åpenhet. De fleste modeller fungerer som "svarte bokser" for ledere, uten å forklare hvordan de kommer frem til konklusjonene sine. Dette fører til problemer: Algoritmer kan forsterke stereotypier og til og med skape humoristiske situasjoner, som i tilfellet med Microsofts chatbot, som raskt ble til et giftig kommunikasjonsverktøy [154].

I Deep Thinking reflekterer Garry Kasparov, tidligere verdensmester i sjakk, over nederlaget mot IBM Big Blue-datamaskinen [155]. Han argumenterer for at den virkelige verdien av AI ikke ligger i å kopiere menneskelig intelligens, men i å utfylle våre evner. Kunstig intelligens bør utføre oppgaver der mennesker er svake, mens mennesker bidrar med kreativitet. Datamaskiner har endret den tradisjonelle måten å analysere sjakk på. I stedet for å skape fascinerende historier om partiene, vurderer datasjakkprogrammene hvert trekk upartisk, kun basert på dets faktiske styrke eller svakhet. Kasparov påpeker at menneskets tendens til å se hendelser som sammenhengende historier i stedet for enkeltstående handlinger ofte fører til feil konklusjoner - ikke bare i sjakk, men i livet generelt.

Hvis du planlegger å bruke maskinlæring til prediksjon og analyse, er det derfor viktig å forstå de grunnleggende prinsippene - hvordan algoritmer fungerer og hvordan data behandles - før du begynner å bruke maskinlæringsverktøyene og AI i arbeidet ditt. Den beste måten å komme i gang på er gjennom praktisk erfaring.

Et av de mest praktiske verktøyene for en første innføring i temaet maskinlæring og prediksjon er Jupyter Notebook og det populære klassiske Titanic-datasettet, som gir en visuell innføring i de viktigste metodene

for å analysere data og bygge ML-modeller.

Titanic-datasettet: Hello World i analyseverdenen data og stordata

Et av de mest kjente eksemplene på bruk av ML i dataanalyse er analysen av Titanic-datasettet, som ofte brukes til å studere sannsynligheten for at passasjerene overlever. Å lære seg denne tabellen kan sammenlignes med "Hello World"-programmet når man lærer seg programmeringsspråk.

Ved Titanics forlis i 1912 omkom 1502 av 2224 personer. Titanic-datasettet inneholder ikke bare informasjon om hvorvidt en passasjer overlevde, men også attributter som alder, kjønn, billettklasse og andre parametere. Datasettet er gratis tilgjengelig og kan åpnes og analyseres på ulike offline- og online-plattformer.

Lenke til Titanic-datasettet:

<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

Tidligere i kapittelet "LLM-aktiverte IDE-er og fremtidige endringer i programmering" har vi allerede omtalt Jupyter Notebook - et av de mest populære utviklingsmiljøene for dataanalyse og maskinlæring. Gratis skyanaloger til Jupyter Notebook er plattformene Kaggle og Google Collab, som gjør det mulig å kjøre Python-kode uten å installere programvare og gir gratis tilgang til databehandlingsressurser.

Kaggle er den største plattformen for dataanalyse og maskinlæringskonkurranser med et integrert kodekjøringsmiljø. Per oktober 2023 har Kaggle over 15 millioner brukere [156] fra 194 land.

Last ned og bruk Titanic-datasettet på Kaggle -plattformen (figur 9.2-5) for å lagre datasettet (en kopi av det) og kjøre Python -kode med forhåndsinstallerte biblioteker direkte i en nettleser, uten å måtte installere en dedikert IDE.

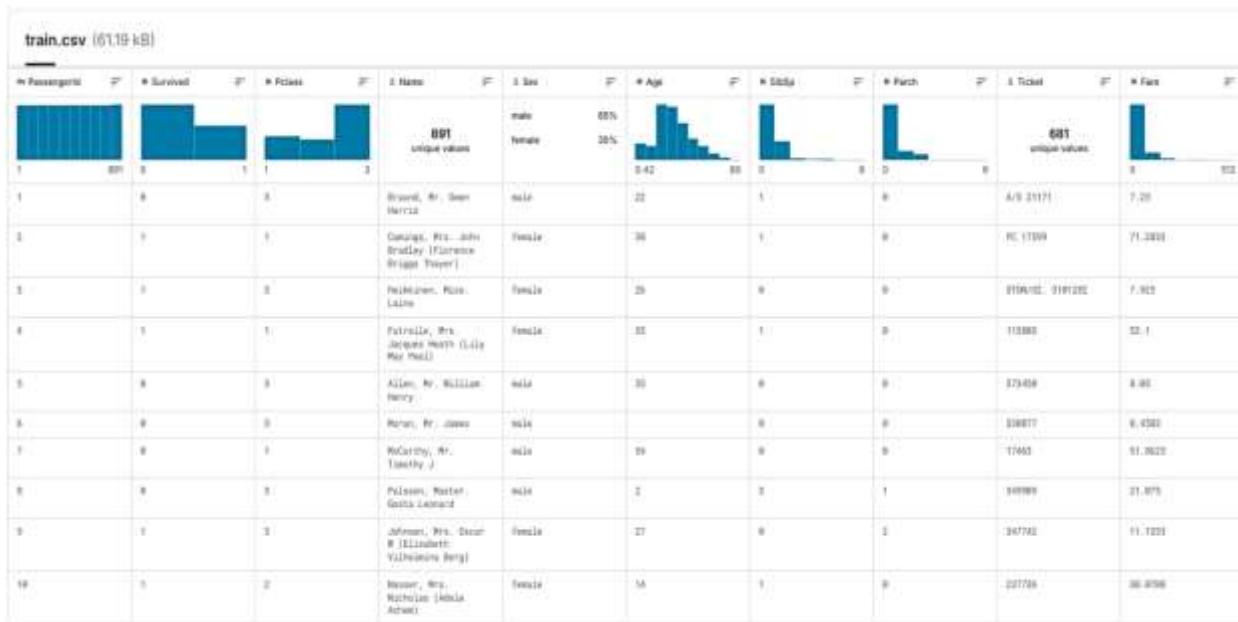


Fig. 9.2-5 Titanic tabellstatistikk - det mest populære treningsdatasettet for dataanalyse og maskinlæring.

Titanic-datasettet inneholder data om de 2224 passasjerene som var om bord på *RMS Titanic* da skipet forliste i 1912. Datasettet presenteres som to separate tabeller, et treningsutvalg (train.csv) og et testutvalg (test.csv), slik at det kan brukes både til å trenere opp modeller og til å evaluere nøyaktigheten deres på nye data.

Treningsdatasettet inneholder både attributter - attributter for passasjerer (alder, kjønn, billettklasse og annet) og informasjon om hvem som overlevde (kolonne med binære verdier "Overlevde"). Treningsdatasettet (figur 9.2-6 - fil train.csv) brukes til å trenere modellen. Testdatasettet (figur 9.2-7 - fil test.csv) inneholder kun passasjerattributter uten informasjon om overlevende (uten en eneste kolonne "Survivor"). Testdatasettet er ment for å teste modellen på nye data og for å evaluere nøyaktigheten.

Dermed har vi nesten identiske attributter for passasjerer i opplærings- og testdatasettene. Den eneste viktige forskjellen er at vi i testdatasettet har en liste over passasjerer som ikke har kolonnen "Survivor" - målvariabelen, som vi ønsker å lære å forutsi ved hjelp av ulike matematiske algoritmer. Og etter at vi har bygget modellen, vil vi kunne sammenligne utdataene fra modellen vår med den virkelige parameteren "Survivor" fra testdatasettet, som vi vil ta hensyn til for å evaluere resultatene.

Tabellens hovedkolonner, passasjerparametere i trenings- og testdatasettet:

- **PassengerId** - unik passasjeridentifikator
- **Overlevde** - 1 hvis passasjen overlevde, 0 hvis død (ikke tilgjengelig i testsettet)
- **Pclass** - billettklasse (1, 2 eller 3)
- **Navn** - passasjerens navn
- **Kjønn** - passasjerens kjønn (mann/kvinne)
- **Alder**
- **SibSp** - antall brødre/søstre eller ektefeller om bord

- **Parch** - antall foreldre eller barn om bord
- **Billett** - billettnummer
- **Billetttakst** - billettpolis
- **Lugar** - lugarnummer (mange data mangler)
- **Embarked** er ombordstigningshavnen (C = Cherbourg, Q = Queenstown, S = Southampton).

For å visualisere manglende data i begge tabellene kan du bruke missingno-biblioteket (fig. 9.2-6, fig. 9.2-7), som viser manglende verdier i form av et histogram, der hvite felt viser manglende data. Denne visualiseringen gjør det mulig å foreta en rask vurdering av datakvaliteten før behandling.

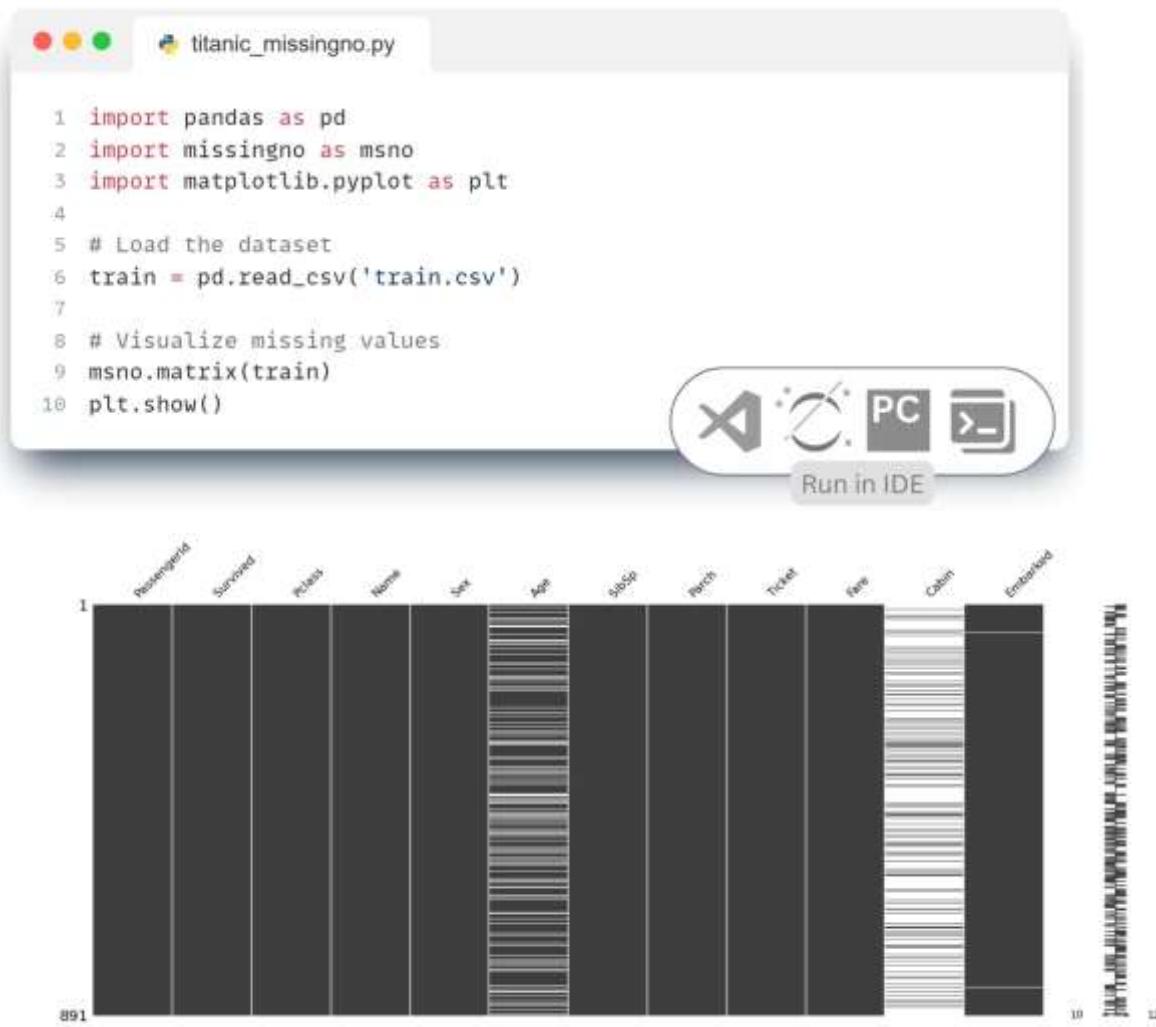
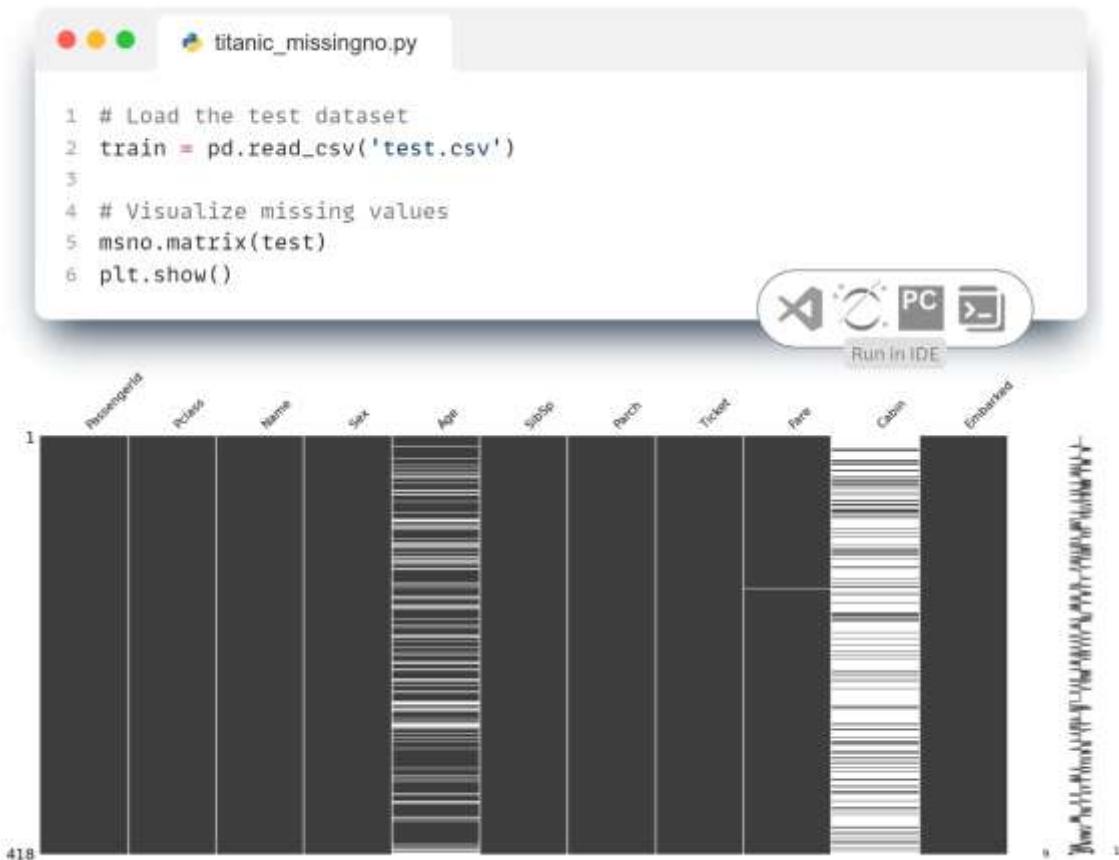


Fig. 9.2-6 Nøn få kodelinjer brukes til å visualisere de manglende dataene i Titanic-treningsdatasettet, der nøkkelparameteren for trening er parameteren "Survived".



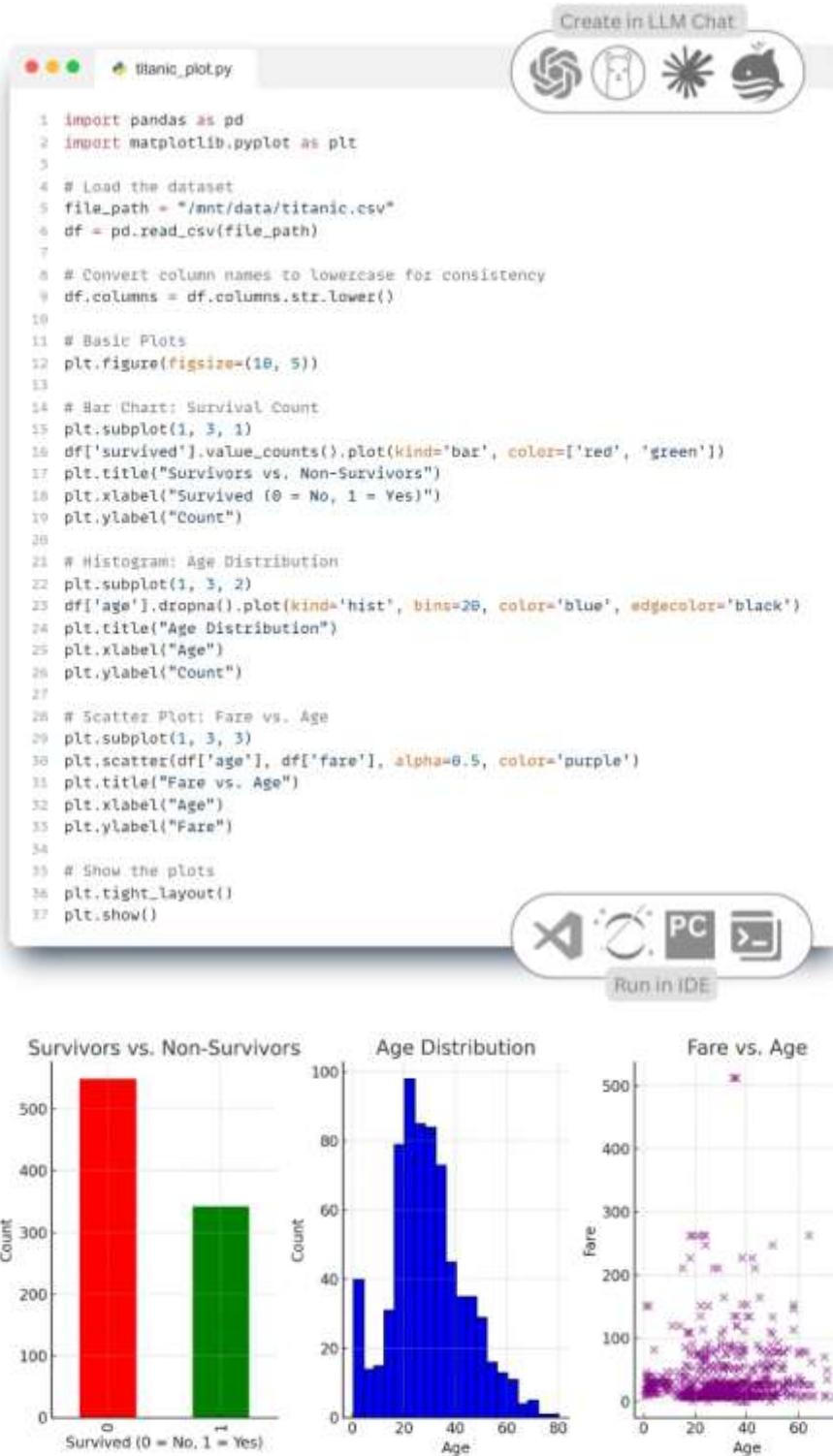
Figur 9.2-7 Visualisering av manglende data i Titanic-testdatasettet, som kun inneholder passasjerkarakteristikker uten informasjon.

Før du formulerer hypoteser og prediksjoner basert på datasettet, kan visuelle analyser bidra til å identifisere viktige mønstre i dataene, vurdere datakvaliteten og identifisere mulige avhengigheter. Det finnes mange visualiseringsteknikker som kan hjelpe deg med å forstå Titanic-datasettet bedre. Du kan bruke fordelingsplotter for å analysere aldersgrupper blant passasjerene, overlevelsediagrammer etter kjønn og klasse og matriser med manglende data for å vurdere kvaliteten på informasjonen og forstå dataene.

- 🔗 La oss be LLM om å hjelpe oss med å visualisere dataene fra Titanic-datasettet ved å sende følgende tekstforespørsel til en hvilken som helst LLM-modell (CHATGP, LLaMa, Mistral, DeepSeek, Grok, Claude, QWEN eller en annen):

Vis noen enkle grafer for Titanic-datasettet. Last ned datasettet selv og vis ↗

- 🔗 LLM-respons i form av ferdig kode grafer som visualiserer datasettparametrene:



Figur 9.2-8 LLM hjelper deg med å få en umiddelbar visualisering av datasettdataene dine.

Datavisualisering er et viktig skritt for å forberede datasettet for den påfølgende konstruksjonen av en

maskinlæringsmodell, som man bare kan få tilgang til ved å forstå dataene.

Maskinlæring i aksjon: fra Titanic-passasjerer til prosjektledelse

Hovedhypotesen som brukes for å utforske maskinlæringsrammeverket basert på Titanic-datasettet, er at visse grupper av passasjerer hadde større sjanse for å overleve.

Den lille tabellen over Titanic-passasjerer har blitt populær over hele verden, og millioner av mennesker bruker den til trening, eksperimentering og modelltesting for å finne ut hvilke algoritmer og hypoteser som vil bygge den mest nøyaktige modellen for prediksjon av overlevelse basert på treningsdatasettet for Titanic-passasjerer.

Titanic-datasettet er attraktivt fordi det er så kompakt: Med flere hundre rader og tolv kolonner (fig. 9.2-6) gir det rikelig med muligheter for analyse. Datasettet er, relativt enkelt sagt, et klassisk eksempel på en binær klassifiseringsløsning, der målet med problemet - overlevelse - er uttrykt i det praktiske formatet 0 eller 1.

John Wheeler i "It from Bit" [7] argumenterer han for at universet er basert på binære valg. På samme måte er en virksomhet som drives av mennesker som består av molekyler, faktisk bygget på en rekke binære valg.

I tillegg er dataene basert på en virkelig historisk hendelse, noe som gjør dem verdifulle for forskning, i motsetning til kunstig skapte eksempler. Bare på Kaggle-plattformen, en av de største Data Pipeline og ETL, deltok 1 355 998 personer i de Titanic-datasettbaserete utfordringene, og utviklet 53 963 unike Data Pipeline-løsninger [157] (fig. 9.2-9).

Det virker utrolig, men bare 1000 linjer med data om Titanic-passasjerene med 12 parametere har blitt et felt for millioner av hypoteser, logiske kjeder og unike datapipelines. Fra et lite datasett fødes endeløse innsikter, hypoteser og tolkninger - fra enkle overlevelsmodeller til komplekse ensembler som tar hensyn til skjulte mønstre og komplekse labyrinter av resonnementer.

Machine Learning from Disaster

Submit Prediction

Data Code Models Discussion Leaderboard Rules

Titanic Tutorial 16916
Updated 3y ago Gold ***
29858 comments · Titanic - Machine Learning from Disaster

Titanic competition w/ TensorFlow Decision Forests 1098
Updated 2y ago Gold ***
Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

Titanic Data Science Solutions 10723
Updated 6y ago Gold ***
2590 comments · Titanic - Machine Learning from Disaster

Exploring Survival on the Titanic 3968
Updated 7y ago Gold ***
Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster

Fig. 9.2-9 De fem første løsningene av totalt 53 963 ferdige Pipeline-løsninger med åpen kildekode. Nesten 1,5 millioner mennesker har allerede prøvd å løse dette problemet bare på Kaggle [157].

Hvis selv en så liten tabell kan generere millioner av unike løsninger (fig. 9.2-9), hva kan vi da si om virkelige industrielle konstruksjonsdatasett der parametrene måles i titusener?

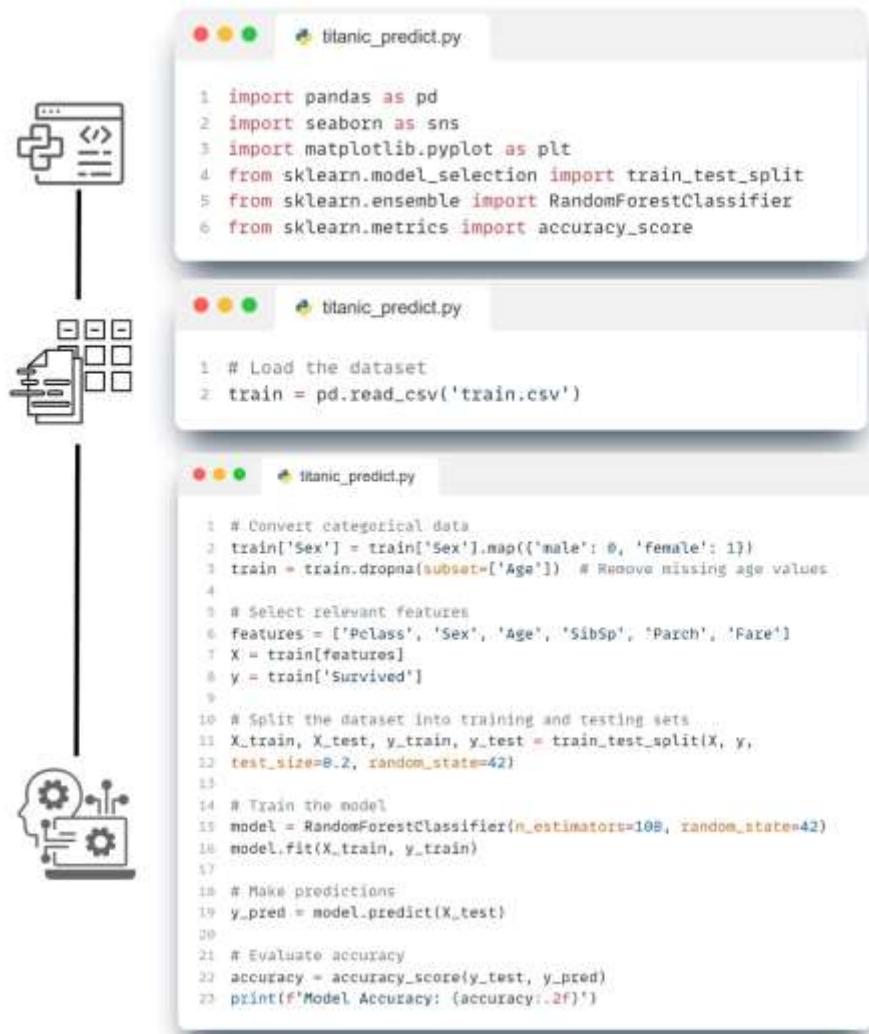
En standard CAD -design av en relativt liten bygning inneholder titusenvis av enheter med tusenvis av parametere - fra geometriske egenskaper til kostnads- og tidsattributter. Tenk deg hvor mange potensielle innsikter, sammenhenger, prediksjoner og styringshypoteser som ligger gjemt i dataene fra alle bedriftens prosjekter som er samlet inn gjennom årene. Historiske prosjektdata er ikke bare et arkiv - det er en organisasjons levende minne, dens digitale fotavtrykk som kan analyseres for å bygge opp et stort antall unike hypoteser.

Og viktigst av alt: Du trenger ikke å vente på at Kaggle -fellesskapet skal fatte interesse for din bedrift eller dine data. Du kan begynne å jobbe med det du har i dag: kjøre analyser på dine egne data, trenre modeller på dine egne data, identifisere gjentakelser, anomalier og mønstre. Der det tidligere krevdes mange års eksperimentering og dyre konsulenttjenester, er det nå nok med initiativ, en LLM, en åpen tilnærming til data og en vilje til å lære.

- ⌚ For å lage en maskinlæringsalgoritme, som kan forutsi overlevelsesraten for passasjerer basert på treningsdatasettet train.csv, ber vi LLM om å løse dette problemet for oss:

Basert på Titanic-passasjerenes treningsdatasett, bygg en maskinlæringsmodell for å forutsi overlevelsesevne ↴

💡 LLMs svar :



```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

1 # Load the dataset
2 train = pd.read_csv('train.csv')

1 # Convert categorical data
2 train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
3 train = train.dropna(subset=['Age']) # Remove missing age values
4
5 # Select relevant features
6 features = ["Pclass", "Sex", "Age", "SibSp", "Parch", "Fare"]
7 X = train[features]
8 y = train['Survived']
9
10 # Split the dataset into training and testing sets
11 X_train, X_test, y_train, y_test = train_test_split(X, y,
12 test_size=0.2, random_state=42)
13
14 # Train the model
15 model = RandomForestClassifier(n_estimators=100, random_state=42)
16 model.fit(X_train, y_train)
17
18 # Make predictions
19 y_pred = model.predict(X_test)
20
21 # Evaluate accuracy
22 accuracy = accuracy_score(y_test, y_pred)
23 print(f'Model Accuracy: {accuracy:.2f}')

```

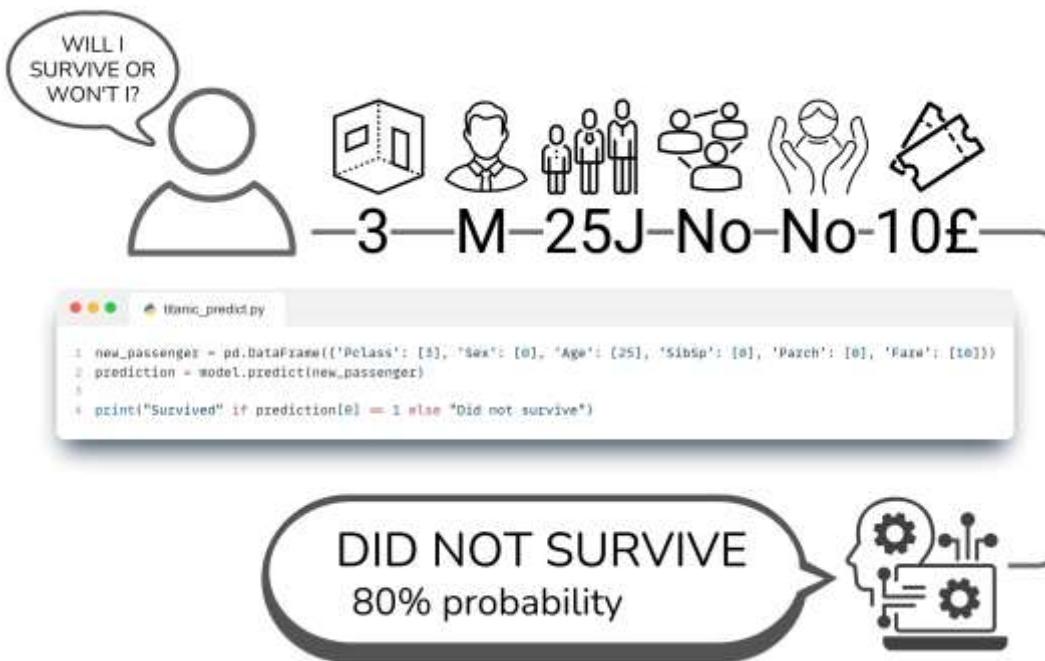
Fig. 9.2-10 LLM laget en prediksjon av overlevende fra Titanic ved hjelp av maskinlæringsalgoritmen Random Forest.

Den resulterende koden fra LLMfig. 9.2-10) laster inn data om Titanic-passasjerene, renser dem, konverterer kategoriske variabler (f.eks. kjønn til numerisk format) og trener modellen ved hjelp av RandomForestClassifier-algoritmen for å forutsi om en passasjer overlevde eller ikke (vi kommer til å snakke mer om populære algoritmer i de neste kapitlene).

Koden deler opp treningsdataene i trenings- og testsett (Kaggles nettsted har allerede opprettet ferdige test.csv (fig. 9.2-7) og train.csv (fig. 9.2-6) for trening, deretter trenes modellen på treningsdataene og testes på testdataene for å se hvor god en bestemt prediksionsmodell er. Etter treningen mates testdataene fra test.csv (med reelle data om de som overlevde eller ikke overlevde) inn i modellen, og den forutsier hvem som overlevde og hvem som ikke gjorde det. I vårt tilfelle er nøyaktigheten til maskinlæringsmodellen vi fikk, ca. 80%, noe som viser at den fanger opp mønstrene ganske godt.

Maskinlæring kan sammenlignes med et barn som prøver å få en rektangulær kloss til å passe inn i et rundt hull. I begynnelsen prøver algoritmen mange tilnæringer, og støter på feil og uoverensstemmelser. Denne prosessen kan virke ineffektiv, men den gir viktig læring: Ved å analysere hver feil forbedrer modellen sine prediksjoner og tar stadig mer nøyaktige beslutninger.

Nå kan denne modellen (Fig. 9.2-10) brukes til å forutsi overlevelsersaten for nye passasjerer, og for eksempel hvis du mater den med passasjerinformasjon ved hjelp av model.predict-funksjonen parametrene: "mann", "3. klasse", "25 år gammel", "ingen slektninger om bord", vil modellen gi en prediksjon - at passasjen med 80% sannsynlighet ikke vil overleve katastrofen hvis han var om bord på Titanic i 1912 (fig. 9.2-11).



Figur 9.2-11 Modellen vi laget ovenfor, kan nå med 80% sannsynlighet forutsi om en ny Titanic-passasjer vil overleve eller ikke.

Modellen for å forutsi Titanic-passasjenes overlevelse illustrerer et mye bredere konsept: Hver dag tar tusenvis av fagfolk i byggebransjen lignende "doble" beslutninger - om liv eller død for en beslutning, et prosjekt, et estimat, et verktøy, fortjeneste eller tap, sikkerhet eller risiko. Som i Titanic-eksempelet, der utfallet var avhengig av faktorer (kjønn, alder, klasse), påvirkes hvert aspekt av beslutningen i byggebransjen av mange av sine egne faktorer og variabler (tabellkolonner): materialkostnader, arbeidernes

kvalifikasjoner, tidsfrister, vær, logistikk, teknisk risiko, kommentarer og hundretusener av andre parametere.

I byggebransjen følger maskinlæring de samme prinsippene som i andre bransjer: Modeller trenes opp på historiske data - fra prosjekter, kontrakter og estimater - for å teste ulike hypoteser og finne de mest effektive løsningene. Denne prosessen er omtrent som å lære opp et barn gjennom prøving og feiling: For hver syklus tilpasser modellene seg og blir mer nøyaktige.

Bruken av akkumulerte data åpner nye horisonter for bygg- og anleggsbransjen. I stedet for tidkrevende manuelle beregninger kan man trenere opp modeller som kan forutsi viktige egenskaper ved fremtidige prosjekter med høy grad av nøyaktighet. På denne måten forvandler prediktiv analyse byggebransjen til et område der man ikke bare kan planlegge, men også forutsi utviklingen med stor sikkerhet.

Forutsigelser og prognosenter basert på historiske data

Dataene som samles inn om selskapets prosjekter, gjør det mulig å bygge modeller som kan forutsi kostnads- og tidskarakteristikker for fremtidige, ennå ikke realiserte objekter - uten tidkrevende manuelle beregninger og sammenligninger. Dette gjør det mulig å fremskynde og forenkle verdsettelsesprosesser betydelig, ikke basert på subjektive antakelser, men på solide matematiske prognosenter.

Tidligere, i bokens fjerde del, har vi sett nærmere på tradisjonelle metoder for estimering av prosjektkostnader, blant annet den ressursbaserte metoden, og vi har også nevnt parametriske metoder og ekspertmetoder. Disse metodene er fortsatt relevante, men i moderne praksis begynner de å bli beriket med verktøy for statistisk analyse og maskinlæring, noe som kan forbedre nøyaktigheten og reproducertbarheten til estimatene betydelig.

Prosessene med manuell og halvautomatisk beregning av priser og tidsattributter vil i fremtiden bli supplert av ML-modeller som er i stand til å analysere historiske data, finne skjulte mønstre og foreslå informerte beslutninger.

Nye data og scenarier vil bli generert automatisk ut fra allerede tilgjengelig informasjon - på samme måte som språkmodeller (LLM) skaper tekster, bilder og kode basert på data som er samlet inn i årenes løp fra åpne kilder [158].

På samme måte som mennesker i dag baserer seg på erfaring, intuisjon og intern statistikk for å vurdere fremtidige hendelser, vil fremtiden for byggeprosjekter i de kommende årene i økende grad bli bestemt av en kombinasjon av akkumulert kunnskap og matematiske maskinlæringsmodeller.

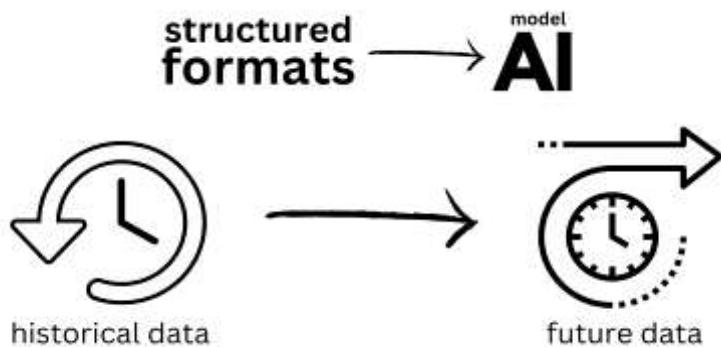


Fig. 9.2-12 Kvalitative og strukturerte historiske bedriftsdata er det materialet som maskinlæringsmodellene og prediksjonene bygger på.

Et enkelt eksempel er å forutsi prisen på et hus basert på areal, tomtestørrelse, antall rom og geografisk beliggenhet. En tilnærming er å bygge en klassisk modell som analyserer disse parameterne og beregner forventet pris (figur 9.2-13). Denne tilnærmingen krever en presis og kjent formel på forhånd, noe som er praktisk talt umulig i praksis.

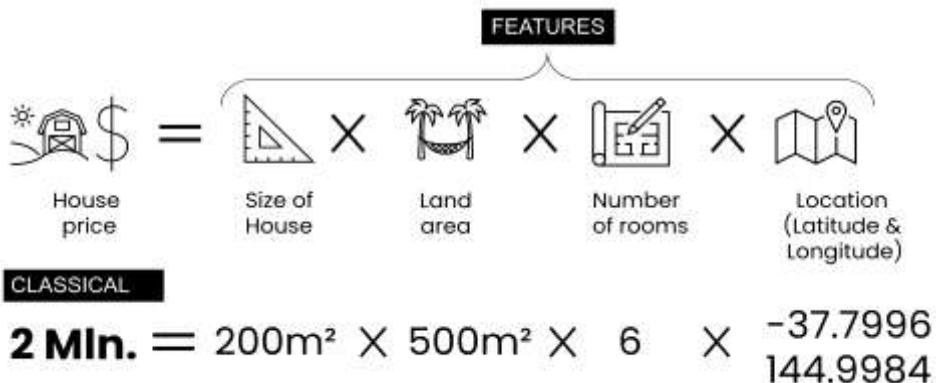


Fig. 9.2-13 En klassisk algoritme kan brukes til å estimere verdien av et hus med en fast formel som skal finnes.

Maskinlæring eliminerer den manuelle letingen etter formler og erstatter dem med opplærte algoritmer som uavhengig av hverandre identifiserer avhengigheter som er mange ganger mer nøyaktige enn forhåndsbestemte ligninger. Alternativt kan vi lage en maskinlæringsalgoritme, som genererer en modell basert på en forhåndsforståelse av problemet og historiske data som kan være ufullstendige (figur 9.2-14).

Med prising som eksempel gjør maskinlæring det mulig å lage ulike typer matematiske modeller som ikke krever kunnskap om den eksakte mekanismen for kostnadsdannelse. Modellen "lærer" av data fra tidligere prosjekter, og tilpasser seg reelle mønstre mellom bygningsparametere, kostnader og tidsfrister.

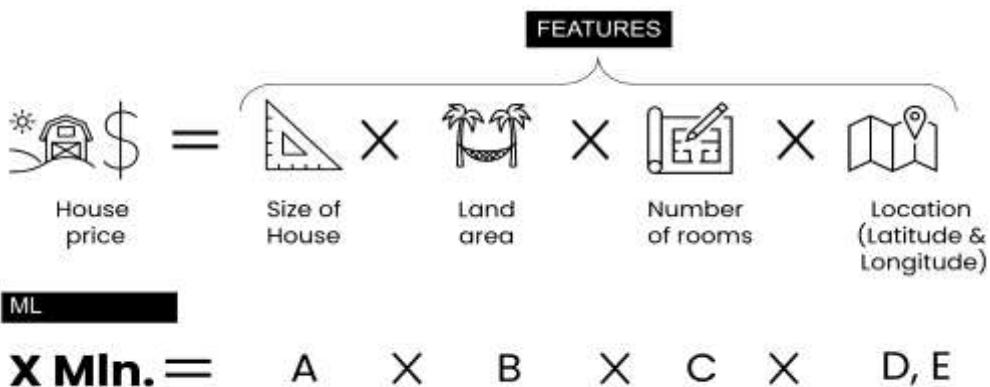
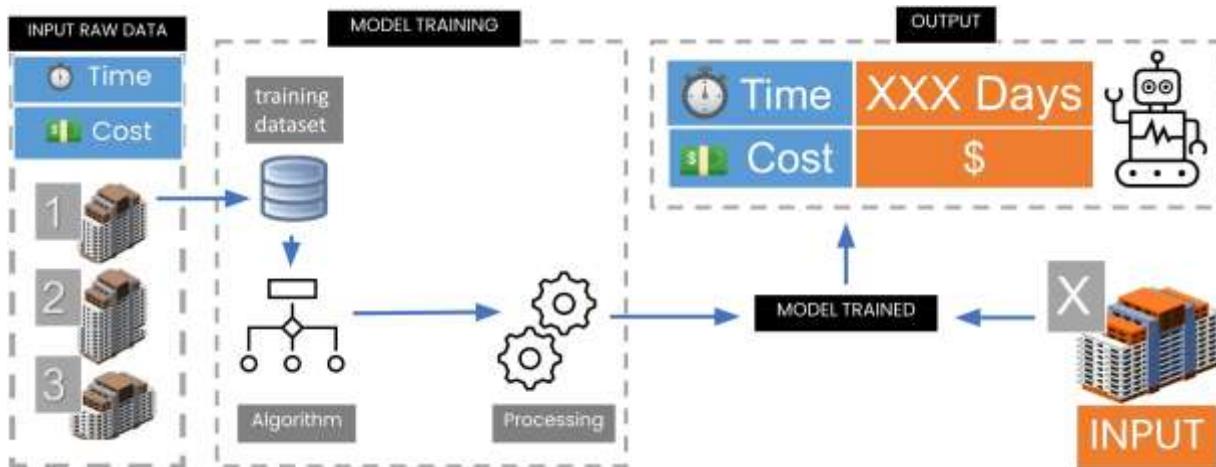


Fig. 9.2-14 I motsetning til klassisk formelbasert estimering, er maskinlæringsalgoritmen trent på historiske data.

I forbindelse med veiledet maskinlæring inneholder hvert prosjekt i opplæringsdatasettet både inndataattributter (f.eks. kostnad og tid for lignende bygninger) og forventede utdataverdier (f.eks. kostnad eller tid). Et lignende datasett brukes til å opprette og tilpasse en maskinlæringsmodell (figur 9.2-15). Jo større datasettet er, og jo bedre kvaliteten på dataene er, desto mer nøyaktig blir modellen og prediksionsresultatene.



Figur 9.2-15 En ML-modell som er trent opp på kostnads- og tidsdata fra tidligere prosjekter, vil med en viss sannsynlighet bestemme kostnadene og tidsplanen for et nytt prosjekt.

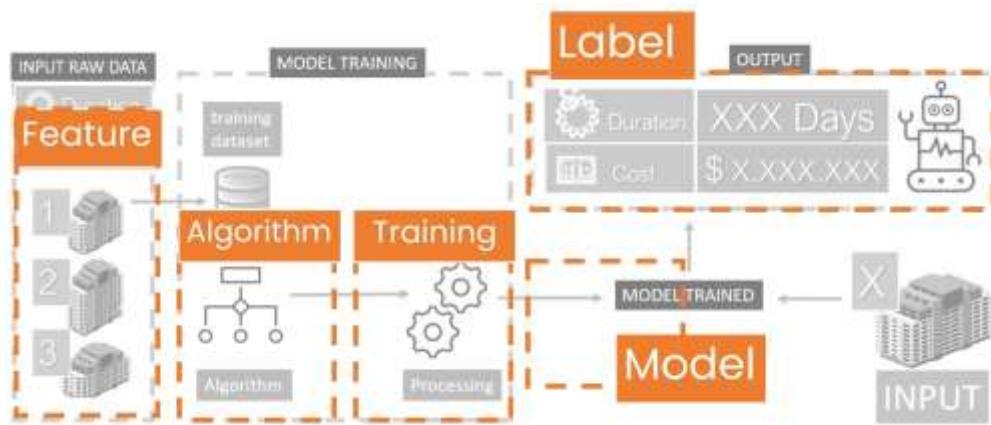
Når modellen er opprettet og opplært til å estimere byggingen av et nytt prosjekt, er det bare å gi modellen nye attributter for det nye prosjektet, så vil modellen gi estimerte resultater basert på tidligere innlærte mønstre med en viss sannsynlighet.

Nøkkelsbegreper innen maskinlæring

Maskinlæring er ikke magi, det er bare matematikk, data og det å finne mønstre. Den har ingen reell intelligens, men er et program som er trent på data for å gjenkjenne mønstre og ta beslutninger uten konstant menneskelig innblanding.

Maskinlæring bruker en rekke nøkkelsbegreper for å beskrive sin struktur (figur 9.2-16):

- **Labels** er målvariabler eller attributter (parameteren "Survivor" i Titanic-datasettet) som modellen skal forutsi. Eksempel: byggekostnad (f.eks. i dollar), varighet av byggearbeidene (f.eks. i måneder).
- **Funksjoner** er uavhengige variabler eller attributter som fungerer som input til modellen. I en prognosemodell brukes de til å forutsi etiketter. Eksempler: tomtestørrelse (i kvadratmeter), antall etasjer i en bygning, bygningens totale gulvareal (i kvadratmeter), geografisk plassering (breddegrad og lengdegrad), type materialer som brukes i byggingen. Antallet egenskaper avgjør også hvor dimensjonale dataene er.
- **En modell** er et sett med ulike hypoteser, hvorav én tilnærmer seg målfunksjonen som skal forutsies eller tilnærmes. Eksempel: maskinlæringsmodellen, som bruker regresjonsanalyseteknikker til å forutsi kostnader og tidspunkt for bygging.
- **Læringsalgoritme Læringsalgoritme** er prosessen med å finne den beste hypotesen i en modell som nøyaktig samsvarer med målfunksjonen ved hjelp av et sett med opplæringsdata. Eksempel: En lineær regresjons-, KNN- eller random forest-algoritme som analyserer kostnads- og byggetidsdata for å identifisere sammenhenger og mønstre.
- **Opplæring** - Under opplæringsprosessen analyserer algoritmen opplæringsdataene og finner mønstre som samsvarer med forholdet mellom inndataattributtene og måletikettene. Resultatet av denne prosessen er en trenet maskinlæringsmodell, klar for prediksjon. Eksempel: en prosess der en algoritme analyserer historiske konstruksjonsdata (kostnader, tid, anleggskarakteristikk) for å lage en prediktiv modell.



Figur 9.2-16 ML bruker etiketter og attributter til å lage modeller som trenes på data ved hjelp av algoritmer for å forutsi utfall.

Maskinlæring eksisterer ikke isolert, men er en del av et bredere økosystem av analytiske disipliner, inkludert statistikk, databaser, datautvinning, mønstergjenkjenning, stordataanalyse og kunstig intelligens. Figur 9.2-17 viser hvordan disse fagfeltene overlapper og utfyller hverandre, og danner et omfattende rammeverk for moderne beslutnings- og automatiseringssystemer.

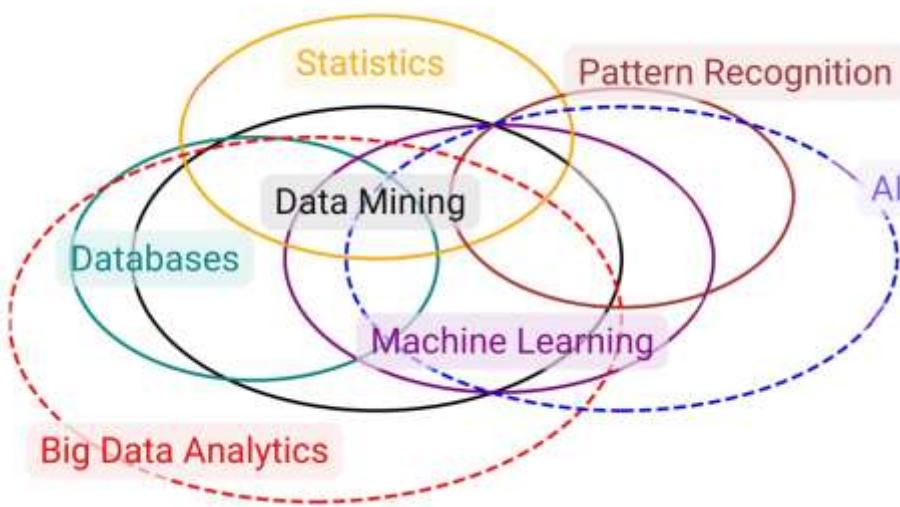


Fig. 9.2-17 Forholdet mellom de ulike områdene innen dataanalyse: statistikk, maskinlæring, kunstig intelligens, stordata, mønstergjenkjenning og data mining.

Hovedmålet med maskinlæring er å gi datamaskiner evnen til automatisk å lære seg kunnskap uten menneskelig inngripen eller assistanse og tilpasse handlingene sine deretter [159].

I fremtiden vil menneskets rolle kun være å gi maskinen kognitive evner - de vil sette betingelsene, vektene og parameterne, og maskinlæringsmodellen vil gjøre resten.

I neste kapittel skal vi se nærmere på konkrete eksempler på bruk av algoritmen. Vi vil bruke virkelige tabeller og forenklede modeller for å vise hvordan prognosene bygges opp trinn for trinn.



KAPITTEL 9.3.

FORUTSI KOSTNADER OG TIDSFRISTER VED HJELP AV MASKINLÆRING

Et eksempel på bruk av maskinlæring for å finne prosjektkostnader og tidslinjer

Estimering av byggetid og -kostnader er en av de viktigste prosessene i et byggefirma. Tradisjonelt har slike estimerer blitt gjort av eksperter basert på erfaring, oppslagsverk og databaser med forskrifter. Men med den digitale transformasjonen og den økende datatilgjengeligheten er det nå mulig å bruke maskinlæringsmodeller (ML) for å forbedre nøyaktigheten og automatiseringen av slike estimerer.

Innføringen av maskinlæring i prosessen med å beregne kostnader og tidsplan for bygging gir ikke bare mer effektiv planlegging, men blir også et utgangspunkt for å integrere intelligente modeller i andre forretningsprosesser - fra risikostyring til optimalisering av logistikk og innkjøp.

Det er viktig å raskt kunne fastslå hvor lang tid det vil ta å bygge et prosjekt, og hva totalkostnaden vil bli. Disse spørsmålene om prosjekttid og -kostnader har tradisjonelt vært sentrale for både byggherrer og byggefirmaer helt siden byggebransjens begynnelse.

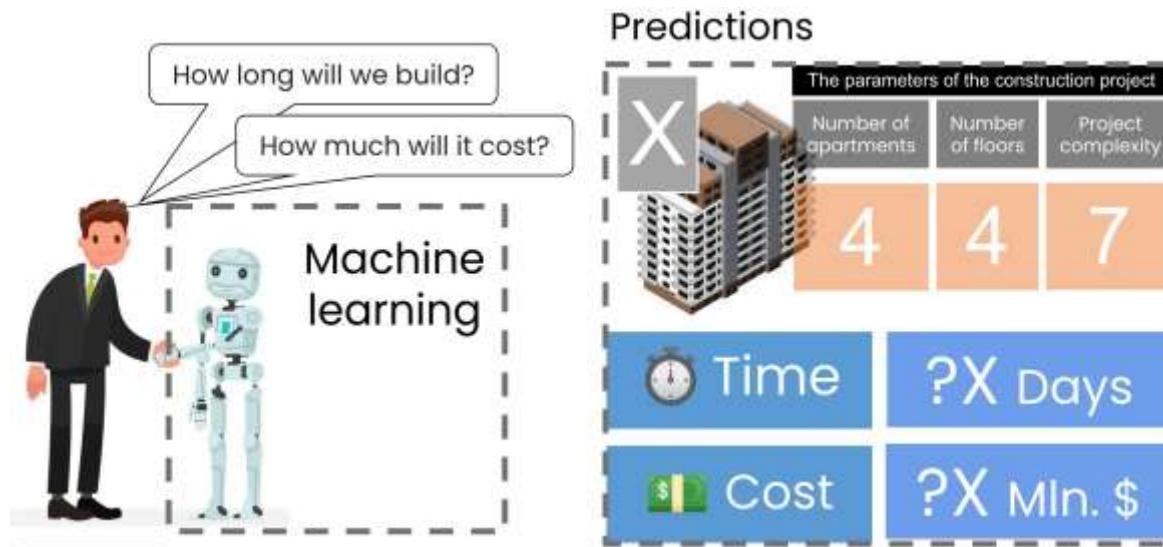


Fig. 9.3-1 I byggeprosjekter er hastighet og kvalitet på estimering av byggetid og -kostnader viktige suksessfaktorer.

I det følgende eksempelet vil nøkkeldata fra tidligere prosjekter bli hentet ut og brukt til å utvikle en maskinlæringsmodell, som vi kan bruke til å estimere kostnader og tidsbruk for nye byggeprosjekter med nye parametere (figur 9.3-1).

Vi tar for oss tre prosjekter med tre nøkkelattributter: antall leiligheter (der 100 leiligheter tilsvarer tallet 10 for å gjøre det lettere å visualisere), antall etasjer og et betinget mål på kompleksiteten i bygningen på en skala fra 1 til 10, der 10 er den høyeste kompleksitetsscoren. Innenfor maskinlæring kalles prosessen med å konvertere og forenkle verdier som 100 til 10 eller 50 til 5 for "normalisering".

Normalisering innen maskinlæring er en prosess der ulike numeriske data bringes til en felles skala for å gjøre det enklere å behandle og analysere dem. Denne prosessen er spesielt viktig når dataene har ulike skalaer og måleenheter.

Anta at det første prosjektet (figur 9.3-2) hadde 50 leiligheter (etter normalisering, 5), 7 etasjer og en kompleksitetsscore på 2, noe som innebar en relativt enkel konstruksjon. Det andre prosjektet hadde allerede 80 leiligheter, 9 etasjer og en relativt kompleks utforming. Under disse forutsetningene tok byggingen av det første og det andre leilighetskomplekset henholdsvis 270 og 330 dager, og den totale prosjektkostnaden var henholdsvis 4,5 og 5,8 millioner dollar.

The figure consists of two tables. The left table, titled 'The parameters of the construction project', shows data for four projects (1, 2, 3, X) across three columns: Number of apartment (5, 8, 3, 4), Number of floors (7, 9, 5, 4), and Project complexity (2, 6, 3, 7). The right table, titled 'The key parameters of the project', shows Time (270, 330, 230, ?X) and Cost (\$ 4.502.000, \$ 5.750.000, \$ 3.262.000, \$?X. XXX.XXX) for the same four projects. Project X is highlighted with orange rows.

Construction project		The parameters of the construction project			The key parameters of the project	
		Number of apartment	Number of floors	Project complexity	Time	Cost
1		5	7	2	270	\$ 4.502.000
2		8	9	6	330	\$ 5.750.000
3		3	5	3	230	\$ 3.262.000
X		4	4	7	?X	\$?X. XXX.XXX

Figur 9.3-2 Et eksempel på et sett med tidligere prosjekter som skal brukes til å estimere tid og kostnader for et fremtidig prosjekt X.

Når vi bygger en maskinlæringsmodell for slike data, er hovedoppgaven å identifisere kritiske attributter (eller etiketter) for prediksjon, i dette tilfellet byggetid og -kostnad. Med et lite datasett vil vi bruke informasjon om tidligere byggeprosjekter til å planlegge nye prosjekter: Ved hjelp av maskinlæringsalgoritmer må vi forutsi kostnadene og byggetiden for et nytt prosjekt X basert på gitte attributter for det nye prosjektet, for eksempel 40 leiligheter, 4 etasjer og en relativt høy prosjektkompleksitet på 7 (fig. 9.3-2). I den virkelige verden kan antallet inngangsparametere være mye større, fra flere titalls til hundrevis av faktorer. Det kan dreie seg om: type byggematerialer, klimasone, entreprenørens kvalifikasjonsnivå, tilgang til offentlige tjenester, type fundament, årstid for oppstart av arbeidene, formannens kommentarer osv.

For å lage en prediktiv maskinlæringsmodell må vi velge en algoritme for å lage den. En algoritme i maskinlæring er som en matematisk oppskrift som lærer datamaskinen hvordan den skal gjøre prediksjoner (blande i riktig rekkefølge av parametere) eller ta beslutninger basert på data.

For å analysere data om tidligere byggeprosjekter og forutsi tid og kostnader for fremtidige prosjekter (figur

9.3-2), kan en populær maskinlæringsalgoritme brukes:

- **Lineær regresjon (Lineær regresjon):** Denne algoritmen prøver å finne en direkte sammenheng mellom attributter, for eksempel mellom antall etasjer og byggekostnad. Målet med algoritmen er å finne en lineær ligning som best beskriver dette forholdet, noe som gjør det mulig å komme med forutsigelser.
- **Algoritme k-nærmeste naboer (k-NN):** Denne algoritmen sammenligner et nytt prosjekt med tidligere prosjekter som er like i størrelse eller kompleksitet. k-NN klassifiserer dataene basert på hvilke av de k (antall) treningseksemplene som ligger nærmest. I regresjonssammenheng er resultatet gjennomsnittet eller medianen av de k nærmeste naboene.
- **Beslutningstrær:** er en prediktiv modellingsmodell som deler inn data i undergrupper basert på ulike forhold ved hjelp av en trestruktur. Hver node i tree representerer en betingelse eller et spørsmål som fører til ytterligere inndeling av dataene, og hvert blad representerer den endelige prediksjonen eller utfallet. Algoritmen deler dataene inn i mindre grupper basert på ulike egenskaper, for eksempel først etter antall historier, deretter etter kompleksitet og så videre, for å komme med en prediksjon.

La oss ta en titt på maskinlæringsalgoritmer for å estimere kostnadene for et nytt prosjekt ved å bruke to populære algoritmer som eksempler: lineær regresjon og K-nærmeste naboer-algoritmen.

Forutsigelse av prosjektkostnader og -tid ved hjelp av lineær regresjon

Lineær regresjon er en grunnleggende dataanalysealgoritme som forutsier verdien av en variabel basert på en lineær sammenheng med en eller flere andre variabler. Denne modellen forutsetter at det er en direkte lineær sammenheng mellom den avhengige variablen og én eller flere uavhengige variabler, og målet med algoritmen er å finne denne sammenhengen.

Den lineære regresjonens enkelhet og klarhet har gjort den til et populært verktøy på en rekke områder. Når man har å gjøre med én enkelt variabel, handler lineær regresjon om å finne den best passende linjen gjennom datapunktene.

Lineær regresjon finner den beste rette linjen (rød linje) som tilnærmer seg avhengigheten mellom inndatavariablene X og utdatavariablene Y. Denne linjen gjør det mulig å forutsi Y-verdier for nye X-verdier basert på det identifiserte lineære forholdet (figur 9.3-3).

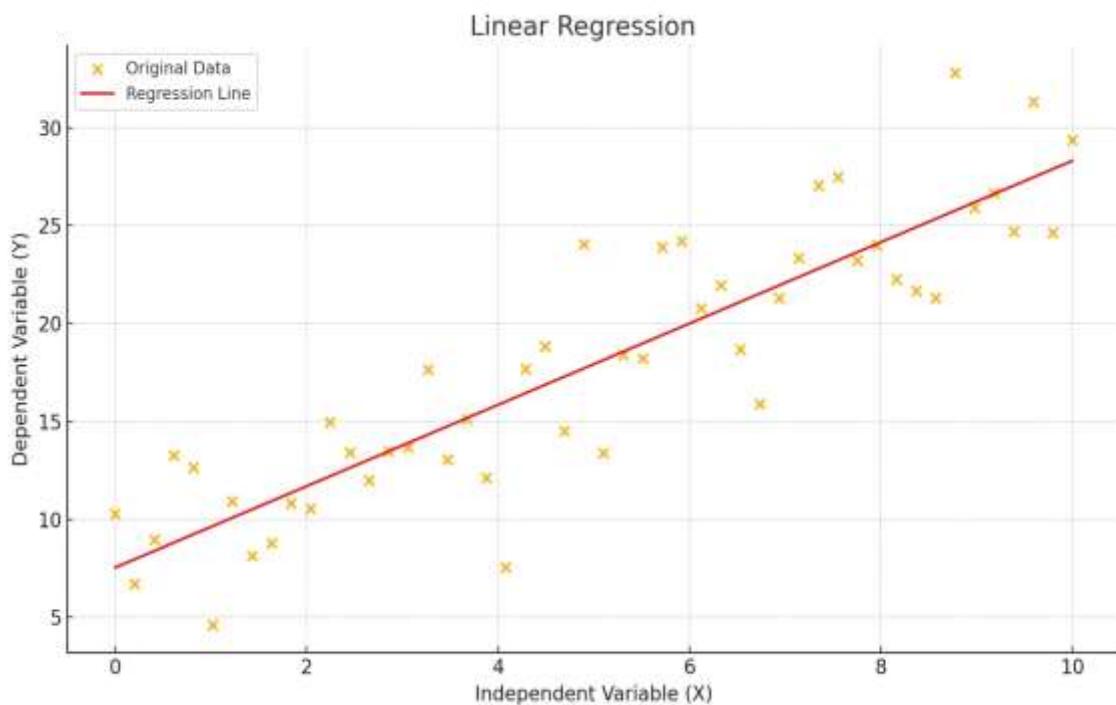


Fig. 9.3-3 Prinsippet for lineær regresjon er å finne den beste rette linjen gjennom treningsverdiene.

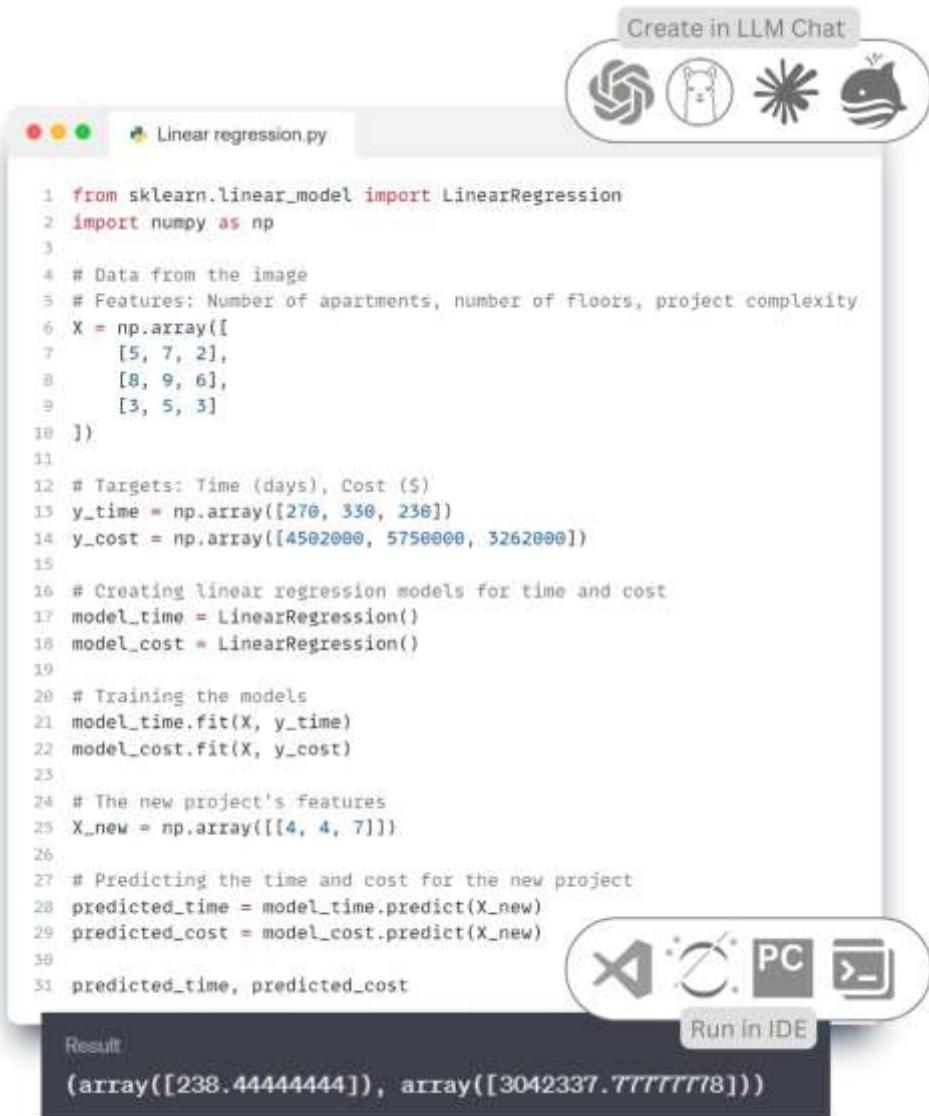
Denne linjen er representert ved en ligning der man ved å legge inn en verdi for den uavhengige variabelen (X) får en predikert verdi for den avhengige variabelen (Y). Denne prosessen gjør det mulig å forutsi Y på en effektiv måte ut fra kjente verdier av X ved hjelp av et lineært forhold mellom dem. Et eksempel på hvordan man kan finne en slik statistisk gjennomsnittlig linje, kan ses i San Francisco Building Permit Data Assessment (figur 9.1-7), der inflasjonen ved hjelp av lineær regresjon ble beregnet for ulike typer anlegg.

La oss laste prosjektdatatabellen (figur 9.3-2 fra forrige kapittel) direkte inn i LLM og be den om å bygge en enkel maskinlæringsmodell for oss.

- Send en tekstforespørsel til LLM chat (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Vi må vise hvordan en enkel maskinlæringsmodell kan konstrueres for å forutsi kostnadene og tiden det tar å realisere et nytt prosjekt X (Fig. 9.3-2 som vedlagt bilde) ↵

■ LLMs svar :



The screenshot shows a user interface for generating code. At the top right is a button labeled "Create in LLM Chat". Below it are four circular icons: a blue hexagonal logo, a white cat face, a yellow starburst, and a blue whale. The main area is a code editor window titled "Linear regression.py". The code is as follows:

```

1 from sklearn.linear_model import LinearRegression
2 import numpy as np
3
4 # Data from the image
5 # Features: Number of apartments, number of floors, project complexity
6 X = np.array([
7     [5, 7, 2],
8     [8, 9, 6],
9     [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

At the bottom of the code editor is a "Run in IDE" button. Below the code editor is a "Result" section containing the output:

```
(array([238.44444444]), array([3042337.77777778]))
```

Figur 9.3-4 LLM valgte lineær regresjon for å lage en maskinlæringsmodell for å forutsi prosjektkostnader og -tid.

LLM gjenkjente automatisk tabellen fra det vedlagte bildet og konverterte dataene fra et visuelt format til en tabellmatrise (fig. 9.3-4 - rad 6). Denne matrisen ble brukt som grunnlag for å lage funksjoner og etiketter som ble brukt til å lage en maskinlæringsmodell (fig. 9.3-4 - 17.-22. rad), der det ble brukt lineær regresjon.

Ved hjelp av en grunnleggende lineær regresjonsmodell som ble trent opp på et "ekstremt lite" datasett, ble det laget prediksjoner for et nytt hypotetisk byggeprosjekt kalt Prosjekt X. I vårt problem kjennetegnes dette prosjektet av 40 leiligheter, 4 etasjer og et kompleksitetsnivå på 7 (figur 9.3-2).

Som forutsagt av en lineær regresjonsmodell basert på et begrenset og lite datasett for det nye prosjekt X (figur 9.3-4 - linje 24-29):

- **Byggetiden** vil være omtrent 238 dager (238.444444)
- **De totale utgiftene** vil være ca. \$ 3 042 338 (3042337,777)

For å utforske prosjektkostnadshypotesen nærmere er det nyttig å eksperimentere med ulike maskinlæringsalgoritmer og -metoder. La oss derfor forutsi de samme kostnads- og tidsverdiene for et nytt prosjekt X basert på et lite sett med historiske data ved hjelp av K-Nearest Neighbours-algoritmen (k-NN).

Prognoser for prosjektkostnader og -tid ved hjelp av K-nærmeste naboalgoritmen (k-NN)

Vi bruker k-Nearest Neighbours-algoritmen (k-NN) som en ekstra prediktor for å estimere kostnadene og varigheten av et nytt prosjekt. K-Nearest Neighbours (k-NN)-algoritmen er en overvåket maskinlæringsmetode (supervised machine learning) for både klassifisering og regresjon. Vi har også tidligere omtalt k-NN-algoritmen i forbindelse med vektordatabasesøk (figur 8.2-2), der den brukes til å finne de nærmeste vektorene (f.eks. tekster, bilder eller tekniske beskrivelser). I denne tilnærmingen representeres hvert prosjekt som et punkt i et flerdimensjonalt rom, der hver dimensjon tilsvarer en bestemt egenskap ved prosjektet.

I vårt tilfelle vil vi representere de tre attributtene til hvert prosjekt som punkter i et tredimensjonalt rom (figur 9.3-5). Vårt kommende prosjekt X vil dermed bli lokalisert i dette rommet med koordinatene ($x=4, y=4, z=7$). Det er verdt å merke seg at i den virkelige verden kan antallet punkter og dimensjonaliteten i rommet være mange ganger større.

K-NN-algoritmen (k-nærmeste naboer) fungerer ved å måle avstanden mellom det ønskede prosjektet X og prosjektene i opplæringsdatabasen. Ved å sammenligne disse avstandene finner algoritmen de prosjektene som ligger nærmest punktet for det nye prosjektet X.

Hvis for eksempel det andre prosjektet ($x=8, y=9, z=6$) fra det opprinnelige datasettet vårt ligger mye lengre unna X (figur 9.3-5) enn de andre prosjektene, kan det utelukkes fra videre analyse. Dermed er det bare de to ($k=2$) nærmeste prosjektene som kan brukes i beregningene, og som gjennomsnittsverdien vil bli fastsatt på grunnlag av.

En slik metode gjør det mulig å vurdere likheter mellom prosjekter ved hjelp av et nabolagssøk, noe som i sin tur bidrar til å trekke konklusjoner om de sannsynlige kostnadene og tidspunktet for et nytt prosjekt basert på lignende prosjekter som har blitt gjennomført tidligere.

k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.

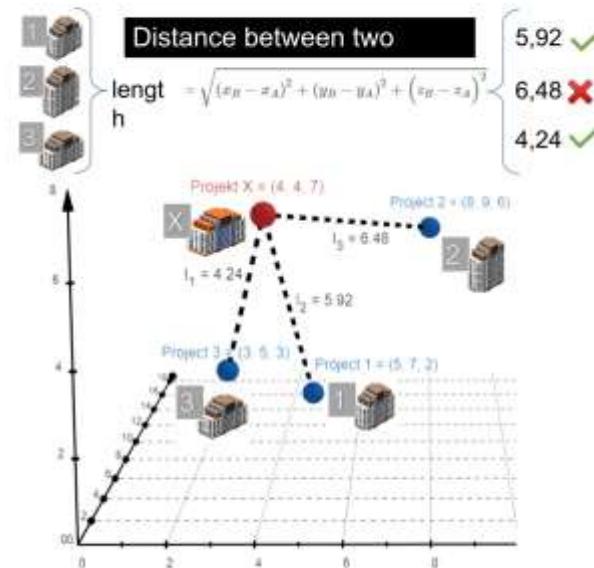
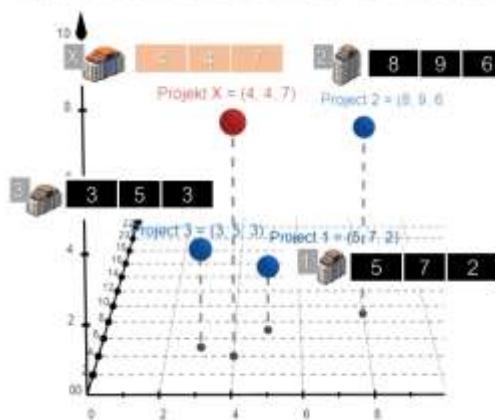


Fig. 9.3-5 I K-NN-algoritmen representeres prosjekter som punkter i et flerdimensjonalt rom, og de nærmeste prosjektene velges ut basert på avstander for likhetsvurdering og prediksjon.

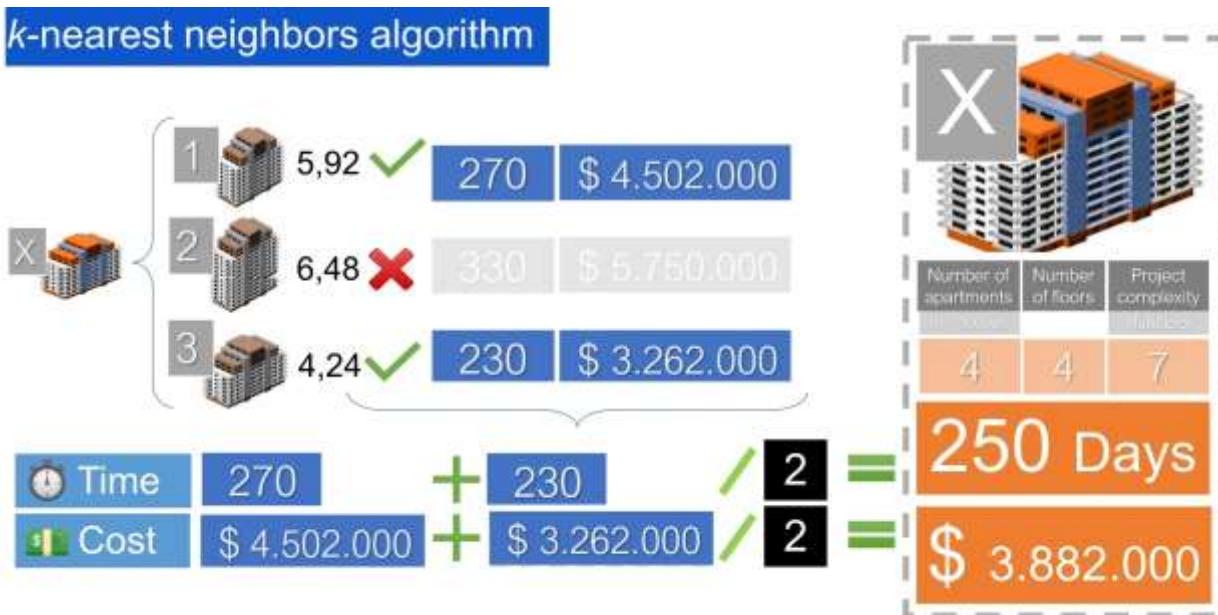
Arbeidet med k-NN omfatter flere viktige trinn:

- **Klargjøring av data:** Trenings- og testdatasett lastes inn først. Treningsdata brukes til å "trene" algoritmen, og testdata brukes til å kontrollere effektiviteten.
- **Valg av parameteren K:** Det velges et tall K som angir hvor mange nærmeste naboer (datapunkter) som skal tas i betrakting i algoritmen. Verdien av "K" er svært viktig fordi den påvirker resultatet.
- Klassifiseringsprosess og regresjon for testdata:
 - **Beregning** av avstander: For hvert element fratestdataene beregnes avstanden til hvert element fra treningsdataene (figur 9.3-5). Ulike metoder for avstandsmåling kan brukes til dette, for eksempel euklidisk avstand (den vanligste metoden), Manhattan-avstand eller Hamming-avstand.
 - **Sortering og valg av K nærmeste naboer:** Etter at avstandene er beregnet, sorteres de, og K punkter som ligger nærmest testpunktet, velges ut.
 - **Bestemme klassen eller verdien til et testpunkt:** Hvis det er en klassifiseringsoppgave, bestemmes klassen til testpunktet basert på den hyppigste klassen blant K utvalgte naboer. Hvis det er en regresjonsoppgave, beregnes gjennomsnittet (eller et annet mål på sentraltendens) av de K naboenes verdier.
- **Fullføring av prosessen:** Når alle testdata er klassifisert eller det er laget prediksjoner for dem, er prosessen fullført.

Algoritmen k-nærmeste naboer (k-NN) er effektiv i mange praktiske anvendelser og er et av de viktigste verktøyene i arsenalet til maskinlæringsespesialister. Algoritmen er populær på grunn av sin enkelhet og effektivitet, særlig i oppgaver der det er lett å tolke forholdet mellom data.

I vårt eksempel ble de to prosjektene (fra vårt lille utvalg) med kortest avstand til prosjekt X identifisert etter å ha brukt K-nærmeste naboer-algoritmen (figur 9.3-5). Basert på disse prosjektene fastsetter algoritmen gjennomsnittet av deres pris og byggetid. Etter analysen (figur 9.3-6) konkluderer algoritmen, ved å beregne gjennomsnittet av de nærmeste naboene, med at prosjekt X vil koste ca. \$ 3 800 000 dollar og ta ca. 250 dager å fullføre.

k-nearest neighbors algorithm



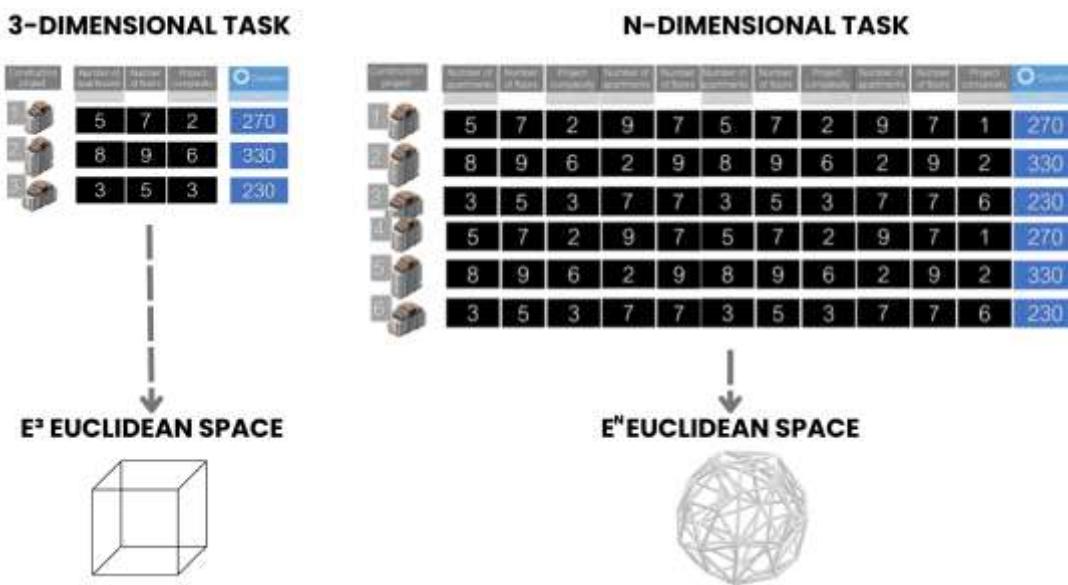
Figur 9.3-6 K-nærmeste naboer-algoritmen bestemmer kostnadene og tidsplanen for prosjekt X ved å analysere de to nærmeste prosjektene i utvalget.

k-Nearest Neighbors (k-NN) -algoritmen er spesielt populær i klassifiserings- og regresjonsoppgaver, for eksempel i anbefalingssystemer, der den brukes til å forestå produkter eller innhold basert på preferanser som ligner på interessene til en bestemt bruker. I tillegg er k-NN mye brukt i medisinsk diagnostikk for å klassifisere typer sykdommer basert på pasientens symptomer, i mønstergjenkjenning og i finanssektoren for å vurdere kredittverdigheten til kunder.

Selv med begrensede data kan maskinlæringsmodeller gi nyttige prediksjoner og forbedre den analytiske komponenten i styringen av byggeprosjekter betydelig. Etter hvert som historiske data utvides og ryddes opp i, er det mulig å gå over til mer sofistikerte modeller - for eksempel ved å ta hensyn til type konstruksjon, beliggenhet, sesong for byggestart og andre faktorer.

I vår forenklede oppgave brukte vi tre attributter for visualisering i 3D-rom, men virkelige prosjekter inneholder i gjennomsnitt hundrevis eller tusenvis av attributter (se datasettet fra kapittelet "An example of CAD-based big data (BIM)", noe som øker dimensjonaliteten i rommet og kompleksiteten ved å

representere prosjekter som vektorer betydelig (figur 9.3-7).



Figur 9.3-7 I det forenkede eksempelet ble det brukt tre attributter for 3D -visualisering, mens virkelige prosjekter har flere.

Ved å bruke ulike algoritmer på det samme datasettet for prosjekt X, som har 40 leiligheter, 4 etasjer og kompleksitetsnivå 7, fikk vi forskjellige predikerte verdier. Den lineære regresjonsalgoritmen predikerte en ferdigstillelsestid på 238 dager og en kostnad på \$ 3 042 338 (figur 9.3-4), mens k-NN-algoritmen predikerte 250 dager og \$ 3 882 000 (figur 9.3-6).

Nøyaktigheten til prediksjonene som oppnås ved hjelp av maskinlæringsmodeller, avhenger direkte av volumet og kvaliteten på inngangadataene. Jo flere prosjekter som er involvert i opplæringen, og jo mer komplette og nøyaktige deres egenskaper (attributter) og resultater (etiketter) er representert, desto større er sannsynligheten for å få pålitelige prediksjoner med minimale feilverdier.

Teknikker for forbehandling av data spiller en viktig rolle i denne prosessen, blant annet

- Normalisering for å bringe funksjonene til en felles skala;
- Deteksjon og eliminering av ekstremverdier, noe som eliminerer modellforvengning;
- Koding av kategoriske attributter for å muliggjøre manipulering av tekstdata;
- Ufylling av manglende verdier øker modellens robusthet.

I tillegg brukes kryssvalideringsmetoder for å vurdere modellens generaliserbarhet og robusthet overfor nye datasett for å avdekke overtildpasning og forbedre påliteligheten til prediksjonen.

Kaos er en orden som skal dechiffreres [160].

- José Saramago, "The Double"

Selv om det virker som om kaoset i oppgavene dine ikke kan beskrives formelt, bør du vite at enhver hendelse i verden og spesielt byggeprosesser er underlagt matematiske lover, som kan trenge støtte fra beregning av verdier, ikke gjennom strenge formler, men ved hjelp av statistikk og historiske data.

Både tradisjonelle kostnadsestimater og maskinlæringsmodeller er uunngåelig befeftet med usikkerhet og potensielle feilkilder. Men hvis det finnes tilstrekkelig med kvalitetsdata, kan maskinlæringsmodeller vise sammenlignbar og noen ganger til og med høyere prediksjonsnøyaktighet enn ekspertvurderinger.

Maskinlæring vil sannsynligvis bli et pålitelig tilleggsverktøy for analyser som kan: avgrense beregninger, foreslå alternative scenarier og identifisere skjulte avhengigheter mellom prosjektparametere. Slike modeller vil ikke gjøre krav på å være universelle, men de vil snart innta en viktig plass i kalkyler og beslutningsprosesser. Maskinlæringssteknologier vil ikke utelukke ingeniører, kalkulatorer og analyticere fra å delta, men vil tvert imot utvide deres muligheter ved å tilby et ekstra synspunkt basert på historiske data.

Hvis maskinlæring integreres på riktig måte i byggefirmaenes forretningsprosesser, har den potensial til å bli et viktig element i ledelsens beslutningsstøttesystem - ikke som en erstatning for mennesker, men som en utvidelse av deres faglige intuisjon og ingeniørlogikk.

Neste steg: fra lagring til analyse og prognoser

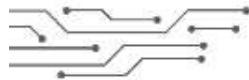
Moderne metoder for å arbeide med data er i ferd med å endre måten beslutninger tas på i byggebransjen. Ved å gå fra intuitive vurderinger til objektiv dataanalyse øker ikke bare nøyaktigheten, men det åpner også for nye muligheter til å optimalisere prosesser. For å oppsummere denne delen er det verdt å fremheve de viktigste praktiske trinnene som vil hjelpe deg med å bruke de diskuterte metodene i det daglige arbeidet:

- Etablering av en bærekraftig infrastruktur for lagring
 - Prøv å kombinere ulike dokumenter og prosjektdata i én enkelt tabellmodell, og samle nøkkelinformasjon i én dataramme for videre analyse
 - Bruk effektive datalagringsformater - for eksempel kolonneformater som Apache Parquet i stedet for CSV eller XLSX - spesielt for datasett som potensielt kan brukes til å trenne opp maskinlæringsmodeller i fremtiden
 - Etablere et system for versjonering av data for å spore endringer gjennom hele prosjektet
- Implementering av analyse- og automatiseringsverktøy
 - Begynn å analysere historiske prosjektdata - via dokumentasjon, modeller og estimater - for å identifisere mønstre, trender og avvik
 - Beherske ETL-prosesser (Extract, Transform, Load) for automatisk innlasting og klargjøring av data
 - Lær hvordan du kan visualisere nøkkeltall ved hjelp av ulike gratis Python-visualiseringsbiblioteker
 - Begynne å bruke statistiske metoder og tilfeldig prøvetaking for å produsere representative og reproducerbare analyseresultater

■ Økende modenhet i arbeidet med data

- Lær noen grunnleggende maskinlæringsalgoritmer med enkle og greie eksempler som Titanic-datasettet
- Analyser nåværende prosesser og identifiser hvor du kan gå fra en rigid årsak-virkning-logikk til statistiske metoder for prognosering og estimering
- Begynn å behandle data som en strategisk ressurs i stedet for et biprodukt: bygg beslutningsprosesser rundt datamodeller i stedet for rundt spesifikke programvareløsninger

Bygg- og anleggsbedrifter som har innsett verdien av data, er på vei inn i en ny utviklingsfase der konkurransefortrinnet ikke avgjøres av mengden ressurser, men av hvor raskt man kan ta beslutninger basert på analyser.



MAKSIMAL BEKVEMMELIGHET MED DEN TRYKTE VERSJONEN

Du holder en gratis digital versjon av **Datadrevet konstruksjon**. For mer praktisk arbeid og rask tilgang til materialet anbefaler vi deg å ta hensyn til den **trykte utgaven**:



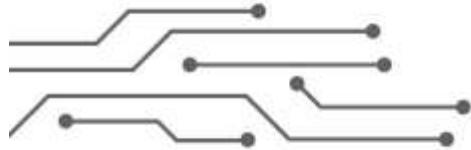
■ **Alltid tilgjengelig:** Boken i trykt format blir et pålitelig arbeidsverktøy, slik at du raskt kan finne og bruke de riktige visualiseringene og diagrammene i enhver arbeidssituasjon

■ **Illustrasjoner av høy kvalitet:** Alle bilder og all grafikk i den trykte utgaven presenteres i maksimal kvalitet

■ **Rask tilgang til informasjon:** enkel navigering, mulighet til å lage notater, bokmerker og arbeide med boken hvor som helst.

Ved å kjøpe den trykte versjonen av boken får du et praktisk verktøy for komfortabelt og effektivt arbeid med informasjon: muligheten til raskt å bruke visuelt materiale i daglige oppgaver, raskt finne de nødvendige ordningene og lage notater. I tillegg støtter kjøpet ditt formidling av åpen kunnskap.

Bestill en trykt versjon av boken på: datadrivenconstruction.io/books



X DEL

BYGGEBRANSJEN I DEN DIGITALE TIDSALDER. MULIGHETER OG UTFORDRINGER

Den siste tiende delen er et omfattende blikk på byggebransjens fremtid i den digitale transformasjonens tidsalder. Her analyseres skiftet fra kausalanalyse til arbeid med stordatakorrelasjoner. Parallelle trekkes mellom utviklingen av kunst og utviklingen av dataarbeid i byggebransjen, og viser hvordan bransjen beveger seg fra detaljstyring til en helhetlig forståelse av prosesser. Konseptet "überisering" av byggebransjen utforskes, der datatransparens og automatiserte beregninger kan endre tradisjonelle forretningsmodeller radikalt, fjerne behovet for mellommenn og redusere mulighetene for spekulasjon. Uløste spørsmål, som den universelle klassifiseringen av elementer, blir diskutert i detalj, slik at byggfirmaene får tid til å tilpasse seg de nye omgivelsene. Delen avsluttes med konkrete anbefalinger for utforming av en strategi for digital transformasjon som inkluderer analyse av sårbarheter og utvidelse av tjenester for å forbli konkurransedyktig i en bransje i endring.

KAPITTEL 10.1.

OVERLEVELSESSTRATEGIER: Å BYGGE KONKURRANSEFORTRINN

Korrelasjoner i stedet for beregninger: Fremtiden for bygganalyse

På grunn av den raske digitaliseringen av informasjon (figur 1.1-5) gjennomgår moderne bygg- og anleggsvirksomhet en grunnleggende endring der data ikke bare blir et verktøy, men en strategisk ressurs som kan endre tradisjonelle tilnæringer til prosjekt- og virksomhetsstyring på en fundamental måte.

I tusenvis av år har byggevirksomhet basert seg på deterministiske metoder - nøyaktige beregninger, detaljering og streng kontroll av parametere. I de første århundrene e.Kr. brukte romerske ingeniører matematiske prinsipper i byggingen av akvedukter og broer. I middelalderen strebet arkitekter etter ideelle proporsjoner i gotiske katedraler, og i industrialiseringen på 1900-tallet ble det utviklet systemer med standardiserte normer og forskrifter som ble grunnlaget for massebygging

I dag går utviklingen fra å lete etter rene årsakssammenhenger til probabilistisk analyse, jakten på korrelasjoner og skjulte mønstre. Bransjen er på vei inn i en ny fase - data er i ferd med å bli en nøkkelressurs, og analyser basert på dem erstatter intuitive og lokalt optimaliserte tilnæringer.



Fig. 10.1-1 Skjult potensial i konstruksjonsdata: Eksisterende beregninger i bedriften er bare toppen av isfjellet som ledelsen kan analysere.

En bedrifts informasjonssystem er som et isfjell (fig. 10.1-1): Bare en liten del av datapotensialet er synlig for bedriftens ledelse, mens den største verdien ligger skjult i dypet. Det er viktig å vurdere dataene ikke bare ut fra hvordan de brukes i dag, men også ut fra hvilke muligheter de vil åpne for i fremtiden. De selskapene som lærer seg å trekke ut skjulte mønstre og skape ny innsikt fra data, som vil skape et varig konkurransefortrinn.

Å finne skjulte mønstre og skape mening ut av data handler ikke bare om å arbeide med tall, men er en kreativ prosess som krever abstrakt tenkning og evnen til å se helheten bak ulike elementer. Slik sett kan utviklingen i arbeidet med data sammenlignes med utviklingen i billedkunsten (figur 10.1-2).

Utviklingen innen bygg og anlegg er bemerkelsesverdig lik utviklingen innen billedkunst. I begge tilfeller har menneskeheten gått fra primitive metoder til sofistikerte visualiserings- og analyseteknologier. I forhistorisk tid brukte menneskene huleristninger og primitive redskaper for å løse hverdaglige oppgaver. I løpet av middelalderen og renessansen økte sofistikeringsnivået i arkitektur og kunst dramatisk. I tidlig middelalder hadde konstruksjonsverktøyene utviklet seg fra en enkel øks til omfattende verktøysett som symboliserer veksten i teknisk kunnskap.

Realismens tidsalder var den første revolusjonen i billedkunsten: Kunstnerne lærte seg å gjengi de minste detaljer for å oppnå maksimal troverdighet. Motstykket til denne perioden var de presise ingeniørteknikkene, de detaljerte tegningene og de strengt regulerte beregningene som ble grunnlaget for designpraksis i århundrer.

Senere endret impresjonismen selve oppfatningen av den kunstneriske virkeligheten: I stedet for å gjengi form bokstavelig, begynte kunstnerne å fange opp stemning, lys og dynamikk, med sikte på å gjenspeile et helhetsinntrykk snarere enn absolutt nøyaktighet. På samme måte beveger maskinlæring innen bygningsanalyse seg bort fra rigide logiske modeller og over til mønstergjenkjenning og probabilistiske mønstre som gjør det mulig å "se" skjulte avhengigheter i data som er utilgjengelige i klassisk analyse. Denne tilnærmingen er et ekko av Bauhaus' ideer om minimalisme og funksjonalitet, der mening (funksjon) er viktigere enn form. Bauhaus forsøkte å fjerne det overflødige, å forlate ornamentikken til fordel for klarhet, utilitarisme og masseappell. Ting skulle være forståelige og nyttige, uten overflødighet - estetikken ble født ut av designens logikk og formål.

Med fotografiets inntog på slutten av 1800-tallet fikk kunsten et nytt verktøy for å fange virkeligheten med enestående nøyaktighet, noe som snudde opp ned på holdningene til billedkunst. På samme måte har den industrielle revolusjonen i det 21. århundret ført til bruk av robotteknologi, lasere, IoT, RFID og konseptet som Connected Construction, der innsamling av individuelle parametere har utviklet seg til en skalerbar, intelligent registrering av hele virkeligheten på byggeplassen.

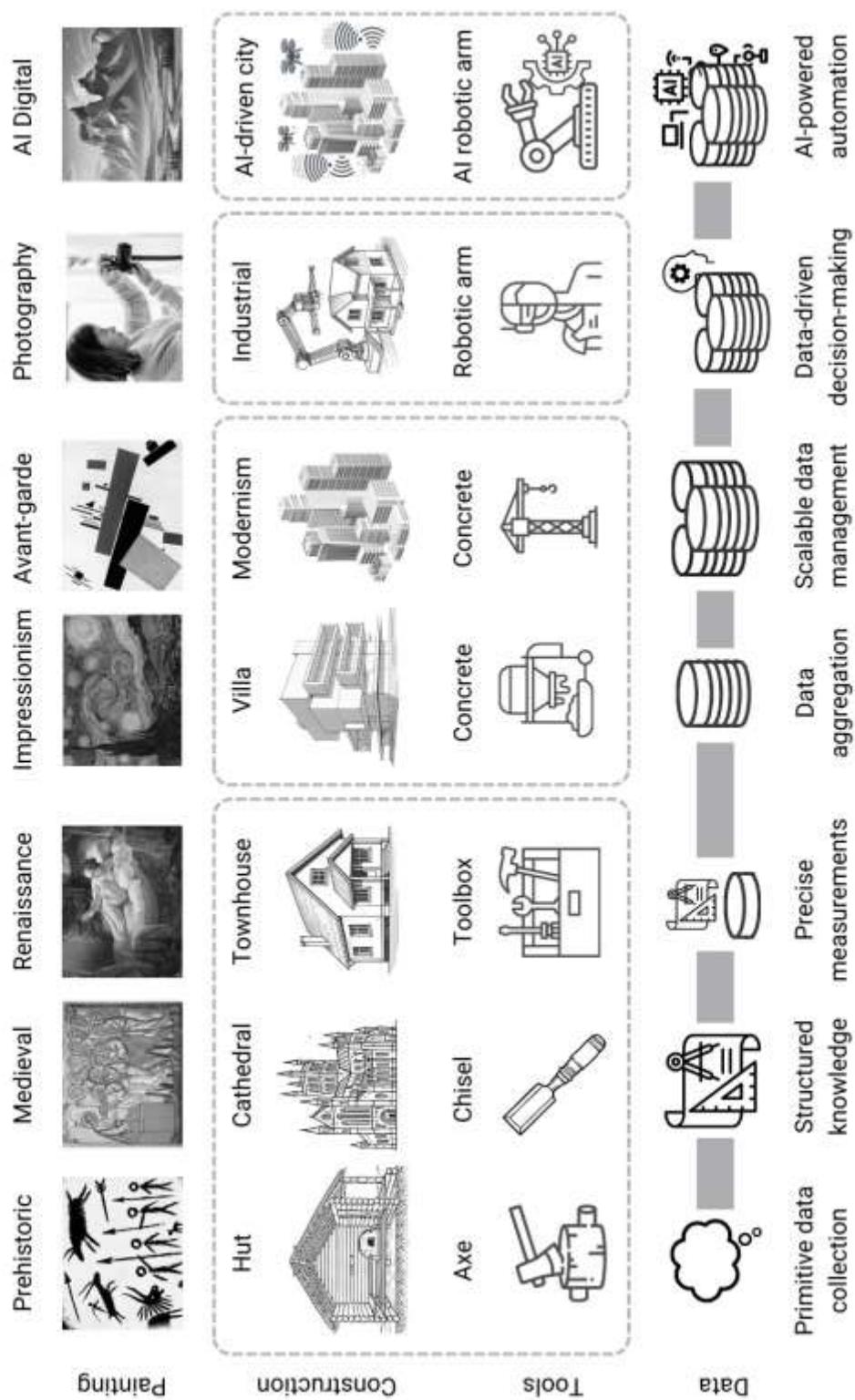


Fig. 10.1-2 Era of Fine Art Evolutions samsvarer med utviklingen måten å jobbe med data på i byggebransjen.

På samme måte som billedkunsten opplever en nytenkning med ankomsten av AI-verktøyene og LLM, opplever byggebransjen i dag et nytt kvantesprang: Intelligent systemer drevet av kunstig intelligens (AI), LLM-chatter, gjør det mulig å forutsi, optimalisere og generere løsninger med minimal menneskelig inngrisen.

Datamaterialets rolle i design og ledelse har endret seg radikalt. Mens kunnskap tidligere ble overført verbalt og empirisk - på samme måte som virkeligheten ble fanget opp av håndmalte bilder frem til 1800-tallet - er dagens fokus på en fullstendig digital fangst av "bildet" av konstruksjonen. Ved hjelp av maskinlæringsalgoritmer omdannes dette digitale bildet til en impresjonistisk representasjon av byggevirkeligheten - ikke en eksakt kopi, men en generalisert, sannsynlighetsbasert forståelse av prosessene.

Vi nærmer oss raskt en tid der prosessene for utforming, konstruksjon og drift av bygninger ikke bare vil bli utvidet, men i stor grad drevet av systemer med kunstig intelligens. På samme måte som moderne digital kunst skapes uten pensel - ved hjelp av tekstlige instruksjoner og generative modeller - vil fremtidens arkitektoniske og tekniske løsninger bli formet av nøkkelspørsmål og parametere som brukeren stiller.

I det 21. århundret er tilgang til data, tolkningen av dem og kvaliteten på analysene blitt uunnværlig for å lykkes med prosjekter. Og verdien av data avgjøres ikke av volumet, men av spesialistenes evne til å analysere, verifisere og omsette dem til handling.

Datadrevet tilnærming i bygg- og anleggsbransjen: et nytt nivå av infrastruktur

I menneskehets historie har hvert eneste teknologiske sprang ført til grunnleggende endringer i økonomi og samfunn. I dag er vi vitne til en ny transformasjonsbølge som i omfang kan sammenlignes med den industrielle revolusjonen på 1800-tallet. Men mens det for hundre år siden var mekaniske krefter og energiteknologi som var den viktigste drivkraften bak endringene, er det nå data og kunstig intelligens.

Maskinlæring, LLM og AI endrer selve essensen i applikasjonene og gjør tradisjonelle programvarestabler (omtalt i bokens andre del) overflødige (fig. 2.2-3). All datalogikk er sentrert i AI-agenter i stedet for i hardkodede forretningsregler (fig. 2.2-4).

I dataalderen er det tradisjonelle synet på applikasjoner i ferd med å bli fundamentalt forandret. Vi beveger oss mot en modell der store, modulære bedriftssystemer uunngåelig må vike plassen for åpne, lette og skreddersydde løsninger.

I fremtiden vil bare den underliggende datastrukturen være igjen, og all interaksjon med den vil skje gjennom agenter som jobber direkte med databasen. Jeg tror virkelig at hele applikasjonsstakken vil forsvinne, fordi det rett og slett ikke er behov for den når kunstig intelligens samhandler direkte med den underliggende databasen. Jeg har brukt hele min karriere på å jobbe i SaaS - bygge selskaper, jobbe i dem, og for å være ærlig ville jeg nok ikke lansert en ny SaaS-virksomhet akkurat nå. Og jeg ville nok ikke investert i SaaS-selskaper akkurat nå. Situasjonen er for usikker. Det betyr ikke at det ikke vil finnes programvareselskaper i fremtiden, men de kommer til å se helt annerledes ut. Fremtidens systemer vil være databaser med forretningslogikk lagt inn i [AI] agenter. Disse agentene vil arbeide med flere datalagre samtidig, og ikke være begrenset til én enkelt database. All logikk vil flyttes inn i AI-laget [46].

- Matthew Berman, administrerende direktør Forward Future

Den viktigste forskjellen ved det nye paradigmet er minimeringen av teknologisk ballast. I stedet for monumentale, komplekse og lukkede programvaresystemer vil vi få fleksible, åpne og raskt tilpassbare moduler som bokstavelig talt "lever" inne i dataflyten (fig. 7.4-1 - Apache Airflow, NiFi). Arkitekturen for framtidens prosessledelse forutsetter bruk av mikroapplikasjoner - kompakte, spesialbygde verktøy som er fundamentalt forskjellige fra massive og lukkede ERP-, PMIS-, CDE, CAFM-systemer. De nye agentene vil være så tilpasningsdyktige, integrerte og virksomhetsspesifikke som mulig (f.eks. Low-Code/No-Code, se 7.4-6).

All forretningslogikken vil gå til disse [AI] agentene, og disse agentene vil utføre CRUD [Create, Read, Update, and Delete]-operasjoner på flere repositorier, noe som betyr at de ikke vil skille mellom hvilken backend som brukes. De oppdaterer flere databaser, og all logikken havner i det såkalte AI -nivået. Og når all logikken ligger i AI-laget, vil folk begynne å bytte ut backends. Vi ser allerede en ganske høy prosentandel av markedsgevinster innen Dynamics-backends og agentbruk, og vi kommer til å bevege oss aggressivt i den retningen, og prøve å samle alt sammen. Enten det er i kundeservice eller på andre områder, for eksempel ikke bare CRM, men også våre økonomi- og driftsløsninger. Fordi folk ønsker mer AI-drevne forretningsapplikasjoner der logikklaget kan drives av AI og AI-agenter. [...]. Noe av det mest spennende for meg er Excel med Python, som kan sammenlignes med GitHub med Copilot. Det vi har gjort, er at nå som du har Excel, er det bare å åpne det, kjøre Copilot og begynne å leke med det. Det er ikke lenger bare å forstå de tilgjengelige tallene - det vil lage en plan på egen hånd. På samme måte som GitHub Copilot Workspace lager en plan og deretter utfører den, er det som en dataanalytiker, som bruker Excel som et verktøy for å visualisere rader og kolonner for analyse. Copilot bruker altså Excel som et verktøy med alle dets muligheter, fordi det kan generere data og har en Python-tolk.

- Satya Nadella, administrerende direktør, Microsoft, intervju med BG2-kanal desember 2024. [28]

Den endringen vi er vitne til i logikken til kontorapplikasjoner - fra modulære, lukkede systemer til AI-agenter som arbeider direkte med åpne data - er bare en del av en mye større prosess. Det handler ikke bare om å

endre grensesnitt eller programvarearkitektur: Endringene vil påvirke de grunnleggende prinsippene for arbeidsorganisering, beslutningstaking og virksomhetsstyring. I byggebransjen vil dette føre til en datadrevet logikk der data blir det sentrale i alle prosesser, fra design til ressursstyring og byggeovervåking.

Neste generasjons digitale kontor: Hvordan AI endrer arbeidsplassen

For nesten hundre år siden opplevde menneskeheden allerede en lignende teknologisk revolusjon. Overgangen fra dampmaskiner til elektriske motorer tok mer enn fire tiår, men førte til en produktivitetsvekst uten sidestykke - først og fremst gjennom desentralisering av energikapasiteten og fleksibiliteten til nye løsninger. Dette skiftet endret ikke bare historiens gang ved å flytte størstedelen av befolkningen fra landsbygda til byene, men la også grunnlaget for den moderne økonomien. Teknologihistorien er en reise fra fysisk arbeid til automatisering og intelligente systemer. På samme måte som traktoren erstattet et utall jordfresere, erstatter moderne digital teknologi tradisjonelle kontorbaserete metoder for byggeledelse (figur 10.1-3). Så sent som på begynnelsen av 1900-tallet arbeidet mesteparten av verdens befolkning for hånd på jorden, inntil mekaniseringen av arbeidet med maskiner og traktorer begynte på 1930-tallet.

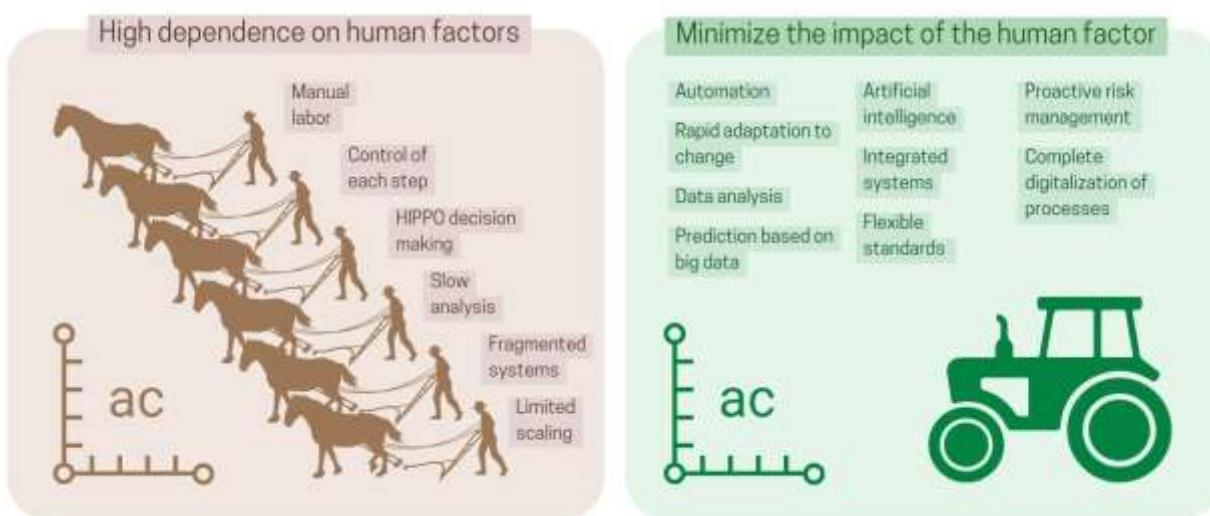


Fig. 10.1-3 Akkurat som traktoren erstattet dusinvis av mennesker på begynnelsen av 1900-tallet, vil maskinlæring erstatte tradisjonelle forretnings- og prosjektstyringsmetoder i det 21. århundret.

På samme måte som menneskeheden for hundre år siden gikk fra å dyrke individuelle jordstykker med primitive redskaper til å drive jordbruk i stor skala ved hjelp av maskiner, går vi i dag fra å behandle ulike "siloer" med informasjon til å arbeide med store datamengder ved hjelp av kraftige "traktorer" - ETL-pipeline og algoritmer for kunstig intelligens.

Vi står foran et lignende sprang - men på det digitale planet: fra tradisjonell, manuell virksomhetsstyring til datadrevne modeller.

Veien til en fullverdig datadrevet arkitektur vil kreve tid, investeringer og organisatorisk innsats. Men denne veien åpner ikke bare for gradvis forbedringer, men også for et kvalitativt sprang mot større effektivitet, åpenhet og håndterbarhet i byggeprosessene. Alt dette forutsetter en systematisk innføring av digitale verktøy og at man går bort fra utdatert forretningspraksis.

Oppgaveparametrisering, ETL, LLM, IoT-komponenter, RFID, tokenisering, **stordata** og maskinlæring vil forvandle tradisjonell bygging til **datadrevet bygging**, der hver eneste detalj i prosjektet og byggevirksomheten vil bli styrt og optimalisert av data

Før tok det tusenvis av arbeidstimer å analysere informasjon. Nå utføres disse oppgavene av algoritmer og LLM-er som gjør ulike datasett om til strategiske kilder ved hjelp av ledetekster. I teknologiverdenen skjer det samme som i landbruket: Vi går fra hakken til automatisert landbruksvirksomhet. Det samme gjelder kontorarbeidet i byggebransjen - fra Excel-filer og manuell oppsummering - til et intelligent system der data samles inn, renses, struktureres og gjøres om til innsikt.

Allerede i dag bør bedrifter begynne å "dyrke" informasjonsfeltene ved hjelp av kvalitetsdatainnsamling og informasjonsstrukturering, "gjødsle" dem med rengjørings- og normaliseringsverktøy og deretter "høste" dem i form av prediktive analyser og automatiserte løsninger. Hvis en moderne bonde med en maskin kan erstatte hundre jordbearbeidere, vil intelligente algoritmer kunne fjerne rutinen fra de ansatte og overføre dem til rollen som strategiske forvaltere av informasjonsflyten.

Det er imidlertid viktig å forstå at det å skape en virkelig datadrevet -organisasjon ikke er en rask prosess. Det er en langsgiktig strategisk retning, på samme måte som å skape et nytt område for å dyrke en ny skog (fig. 1.2-5) av systemer, der hvert "tre" i dette økosystemet er en separat prosess, kompetanse eller et verktøy som det tar tid å utvikle og utvikle. Som i en ekte skog avhenger suksessen ikke bare av kvaliteten på plantematerialet (teknologi), men også av jordsmonn (bedriftskultur), klima (forretningsmiljø) og pleie (systemtilnærming).

Bedrifter vil ikke lenger kunne basere seg utelukkende på lukkede, ferdigutviklede løsninger. I motsetning til tidligere stadier i den teknologiske utviklingen vil den nåværende overgangen - til åpen datatilgang, kunstig intelligens og åpen kildekode - neppe støttes av de store leverandørene, ettersom den direkte truer deres etablerte forretningsmodeller og kjerneinntektsstrømmer.

Som det fremgår av en studie fra Harvard Business School [40], som allerede er omtalt i kapittelet om den fjerde og femte teknologiske revolusjonen, ville kostnadene ved å lage de mest brukte Open Source-løsningene fra bunnen av for alle selskaper være på rundt 4,15 milliarder dollar. Hvis vi derimot forestiller oss at hvert enkelt selskap skulle utvikle sine egne alternativer uten tilgang til eksisterende Open Source-verktøy, noe som har skjedd de siste tiårene, ville de totale kostnadene for næringslivet kunne komme opp i kolossale 8,8 billioner dollar - dette er prisen på irrasjonell etterspørsel som programvaremarkedet kan verdsettes til.

Den teknologiske utviklingen vil uunngåelig føre til en revurdering av etablerte forretningsmodeller. Mens selskaper tidligere kunne tjene penger på komplekse, ugjennomsiktige prosesser og lukkede data, blir denne tilnærmingen mindre og mindre levedyktig i takt med utviklingen av AI og analytics.

Demokratiseringen av tilgangen til data og verktøy kan føre til at det tradisjonelle markedet for salg av programvare krymper betydelig. Samtidig vil et nytt marked vokse frem - markedet for digital ekspertise, skreddersøm, integrasjon og løsningsdesign. Her vil verdien ikke komme fra lisenssalg, men fra evnen til å bygge fleksible, åpne og tilpasningsdyktige digitale prosesser. På samme måte som elektrifiseringen og traktorenes inntog har skapt nye bransjer, vil anvendelsen av stordata, AI og LLM åpne helt nye horisonter for byggebransjen, noe som ikke bare vil kreve teknologiske investeringer, men også en dyptgripende endring av tankesett, prosesser og organisasjonsstrukturer. Og de selskapene og fagfolkene som innser dette og begynner å handle i dag, vil bli morgendagens ledere.

I en verden der åpne data er i ferd med å bli en stor ressurs, vil tilgjengeligheten av informasjon være en game changer. Investorer, kunder og tilsynsmyndigheter vil i økende grad kreve åpenhet, og maskinlæringsalgoritmer vil automatisk kunne identifisere avvik i estimater, tidslinjer og kostnader. Dette skaper forutsetninger for en ny fase i den digitale transformasjonen, som gradvis fører oss til en "uberisering" av byggebransjen.

Åpne data og Uberisering er en trussel mot eksisterende bygg- og anleggsvirksomhet

Byggebransjen er i ferd med å bli en informasjonsstyringsprosess. Jo mer nøyaktige, gode og fullstendige dataene er, desto mer effektive blir design, beregninger, kostnadsoverslag, oppføring og drift av bygninger. I fremtiden vil den viktigste ressursen ikke være kran, betong og armeringsjern, men evnen til å samle inn, analysere og bruke informasjon.

Byggherrer - investorer og byggherrer som finansierer bygg - vil unngåelig utnytte verdien av åpne data og analyser av historiske data i fremtiden. Dette vil åpne for muligheter til å automatisere beregningen av prosjektets tidslinjer og kostnader, uten å involvere byggefirmaene i kostnadsberegninger, noe som vil bidra til kostnadskontroll og raskere identifisering av overflødige kostnader.

Se for deg en byggeplass der laserskannere, quadrokoptere og fotogrammetrisystemer samler inn nøyaktige sanntidsdata om hvor mye betong som er brukt. Denne informasjonen konverteres automatisk til enkle, flate MESH-modeller med metadata, utenom tungvinte CAD (BIM)-systemer, uten å være avhengig av komplekse geometriske kjerner, ERP eller PMIS. Disse dataene som samles inn fra byggeplassen, overføres sentralt til et enkelt strukturert lager som er tilgjengelig for kunden for uavhengig analyse, der reelle priser fra ulike byggeforretninger lastes opp, og parametere som spenner fra kreditfinansiering til dynamisk skiftende faktorer som værforhold, børsnoteringer for byggematerialer, logistikk tariff og statistiske sesongsvingninger i arbeidskraftpriser. Under slike forhold blir eventuelle avvik mellom prosjekterte og faktiske materialvolumer umiddelbart åpenbare, noe som gjør det umulig å manipulere estimater både på prosjekteringsstadiet og når prosjektet leveres. Resultatet er at gjennomsiktigheten i byggeprosessen ikke oppnås gjennom en hærskare av veiledere og ledere, men gjennom objektive digitale data, noe som minimerer den menneskelige faktoren og muligheten for spekulasjoner.

I fremtiden vil denne typen datakontrollarbeid heller bli utført av dataforvaltere på kundesiden (fig. 1.2-4)

CQMS manager). Dette gjelder spesielt for kalkyler og prosjektkalkyler: Der det tidligere fantes en hel avdeling med kalkulatører, vil det allerede i morgen finnes maskinlæring og prognoseverktøy som setter prisgrenser som byggefirmaene skal passe inn i.

Gitt den fragmenterte karakteren til [byggebransjen], der de fleste systemene og delsystemene leveres av små og mellomstore bedrifter, må den digitale strategien komme fra kunden. Kundene må skape forutsetninger og mekanismer for å frigjøre de digitale mulighetene i leverandørkjeden [20].

- Andrew Davis og Giuliano Denicol, Accenture "Å skape mer verdi gjennom kapitalprosjekter"

En slik åpenhet og gjennomsiktighet i dataene utgjør en trussel mot byggefirmaer, som er vant til å tjene penger på ugnjennomsiktige prosesser og forvirrende rapporter, der spekulasjoner og skjulte kostnader kan skjule seg bak komplekse og lukkede formater og modulære, proprietære dataplattformer. Derfor er det lite sannsynlig at byggefirmaer, i likhet med leverandører som promoterer Open Source-løsninger, er interessert i å implementere åpne data fullt ut i forretningsprosessene sine. Hvis dataene er tilgjengelige og enkle å behandle for kunden, kan de kontrolleres automatisk, noe som eliminerer muligheten for overestimering av volumer og manipulering av estimater.

Ifølge World Economic Forums rapport "Shaping the Future of Construction" (2016) [5] er en av de viktigste utfordringene for bransjen fortsatt kundens passive rolle. Likevel er det kundene som bør ta et større ansvar for utfallet av prosjektene - fra tidlig planlegging til valg av bærekraftige samhandlingsmodeller og resultatoppfølging. Uten prosjekteierne aktive deltagelse er det umulig å få til en systemisk endring av byggebransjen.

Tapet av kontroll over volum- og kostnadsberegninger har allerede forvandlet andre bransjer i løpet av de siste 20 årene, slik at kundene kan gå direkte, uten mellomledd, til stymie sine mål. Digitalisering og datatransparens har endret mange tradisjonelle forretningsmodeller, for eksempel drosjesjåfører med fremveksten av Uber (figur 10.1-4), hotelleiere med fremveksten av Airbnb og detaljister og butikker med fremveksten av Amazon, og banker med fremveksten av neo-banker og desentraliserte fintech-økosystemer, der direkte tilgang informasjon og automatisering av tids- og kostnadsberegninger har redusert mellomleddenes rolle betydelig.

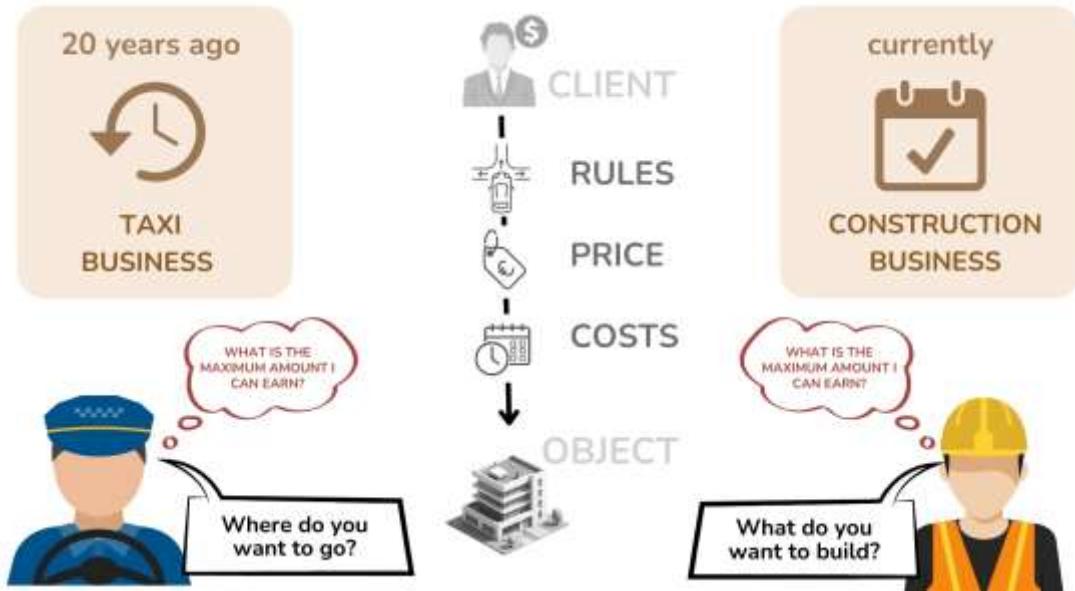


Fig. 10.1-4 Byggebransjen vil møte den samme uberiseringen som drosjesjåfører, hotelleiere og detaljhandlere måtte møte for ti år siden.

Demokratiseringen av tilgangen til data og verktøy for databehandling er uunngåelig, og over tid vil åpne data om alle prosjektkomponenter bli et kundekrav og en ny standard. Derfor vil spørsmålet om å innføre åpne formater og transparente beregninger bli fremmet av investorer, kunder, banker og private equity-fond (private equity) - de som er sluttbrukerne av anleggene som bygges, og som deretter skal driftet dem i flere tiår.

Store investorer, kunder og banker krever allerede åpenhet i byggebransjen. Ifølge Accentures studie "Creating more value through capital projects" (2020) [20] er transparente og pålitelige data i ferd med å bli en avgjørende faktor for investeringsbeslutninger i bygg- og anleggsbransjen. Ekspertene påpeker at pålitelig og effektiv prosjektstyring er umulig uten åpenhet, spesielt i krisetider. I tillegg går eiendomsbesittere og entreprenører i økende grad over til kontrakter som gir incentiver til datadeling og samarbeid om analyser, noe som gjenspeiler de økende kravene fra investorer, banker og tilsynsmyndigheter om ansvarlighet og åpenhet.

Investorens, kundens, bevegelse fra idé til ferdig bygg vil i fremtiden være som å reise på autopilot - uten en sjåfør i form av et byggefirma, og lover å bli uavhengig av spekulasjoner og usikkerhet.

Tiden med åpne data og automatisering vil uunngåelig endre byggebransjen, akkurat som den allerede har gjort i bank-, handels-, landbruks- og logistikkbransjen. I disse bransjene viker mellomleddenes rolle og tradisjonelle måter å gjøre forretninger på for automatisering og robotisering, noe som gjør at det ikke lenger er rom for uberettigede påslag og spekulasjon.

Dataene og prosessene i all menneskelig økonomisk virksomhet er ikke annerledes enn det fagfolk i byggebransjen må forholde seg til. På lang sikt kan byggefirmaer, som i dag dominerer markedet ved å sette standarder for pris og servicekvalitet, miste sin rolle som et viktig mellomledd mellom kunden og byggeprosjektet.

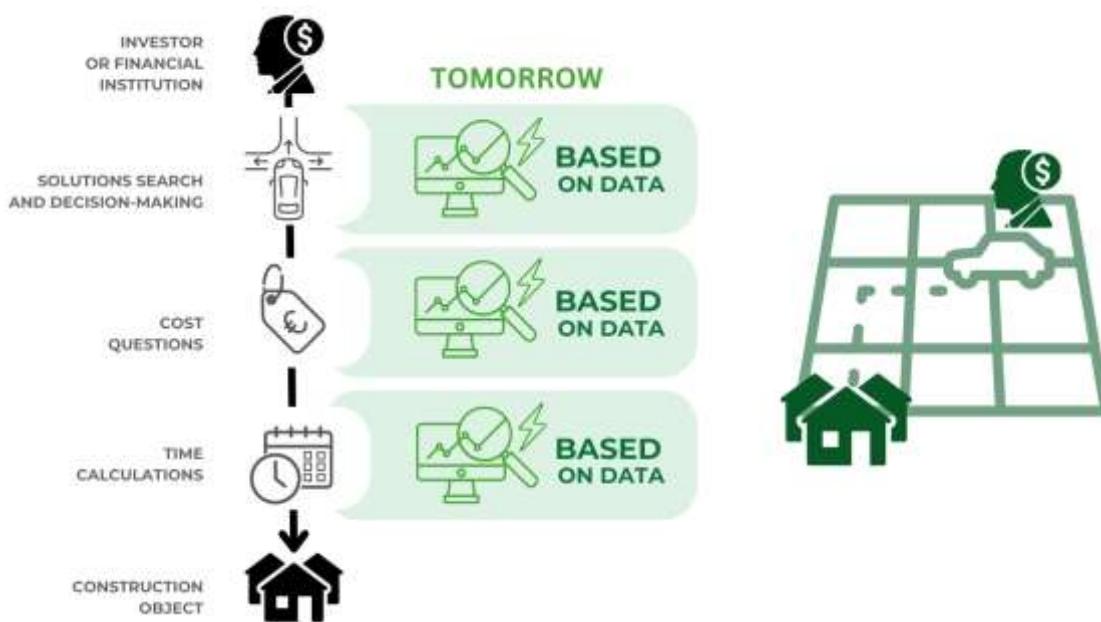
Ulöste problemer med uberisering som siste sjanse til å bruke tiden til transformasjon

Men la oss gå tilbake til realitetene i byggebransjen. Selv om selvkjørende biler, desentraliserte finanssystemer og løsninger basert på kunstig intelligens er på vei inn i enkelte sektorer av økonomien, er en betydelig andel av bygg- og anleggsbedriftene fortsatt papirbaserte organisasjoner der viktige beslutninger i større grad tas på grunnlag av intuisjon og erfaring fra individuelle spesialister.

I dette paradigmet kan et moderne byggefirma sammenlignes med et 20 år gammelt drosjeselskap som kontrollerer ressurser, ruter og leveringstider, og som er ansvarlig for tidspunktet og kostnadene for "reisen" - fra prosjektideen (logistikk- og installasjonsprosessen) til leveransen av prosjektet. På samme måte som GPS (i bygg- og anleggsbransjen IoT, RFID) og maskinlæringsalgoritmer i tids- og kostnadsberegninger en gang forvandlet transportbransjen, har data, algoritmer og AI-agenter potensial til å forandre byggeledelse - fra intuitive vurderinger til prediktive, veiledede modeller. I løpet av de siste 20 årene har mange bransjer - finans, landbruk, detaljhandel og logistikk - gradvis eliminert muligheten til å spekulere på grunn av u gjennomsiktige data. Priser, leveringskostnader eller finansielle transaksjoner beregnes automatisk og statistisk sikkert - på bare noen få sekunder på digitale plattformer.

Fremover må byggefirmaene innse at en demokratisering av tilgangen til data og verktøyene for å analysere dem vil bryte med den tradisjonelle tilnærmingen til estimering av prosjektkostnader og tidsrammer, og fjerne muligheten til å spekulere i u gjennomsiktige volum- og prisdata.

Som å kjøre på en regulert vei uten at sjåføren griper inn, vil fremtidens byggeprosesser i økende grad ligne et "Uberisert" system - med automatisert tids- og kostnadsestimering, transparent ruting av oppgaver og minimal avhengighet av menneskelige faktorer. Dette vil endre selve "reisen" fra idé til realisering - og gjøre den mer forutsigbar, håndterbar og datadrevet.



Figur 10.1-5 Kostnader og "reisetid" under byggingen bestemmes ved hjelp av maskinlæring og statistiske verktøy.

Med den gradvise innføringen av nye forskrifter og krav i nesten alle land i verden, som pålegger overføring av CAD- (BIM)-modeller til byggherrer eller banker som finansierer byggeprosjekter, har byggherren og kunden mulighet til på egen hånd å sikre transparens i kostnads- og omfangskalkyler. Dette er spesielt viktig for store kunder og investorer, som har tilstrekkelig kompetanse og verktøy til å analysere volumene raskt og overvåke markedsprisene. For selskaper som gjennomfører store standardprosjekter - butikker, kontorbygg, boligkomplekser - er slik praksis i ferd med å bli standard.

Etter hvert som informasjonsinnholdet i modellene blir mer komplett og standardisert, forsvinner muligheten for manipulasjon og spekulasjon. Den digitale transformasjonen endrer gradvis spillereglene i byggebransjen, og selskaper som ikke tilpasser seg disse endringene, kan stå overfor alvorlige utfordringer.

Økt konkurranse, teknologiske omveltninger og krympende marginer kan få konsekvenser for bedriftens bærekraft. Med begrenset likviditet vender stadig flere aktører i bransjen seg til automatisering, analyse og data teknologi som en måte å forbedre effektiviteten og prosesstransparense på. Disse verktøyene er i ferd med å bli en viktig ressurs for å bli konkurransedyktig i et økonomisk miljø i endring.

Kanskje vi ikke bør vente på at ytre omstendigheter skal tvinge oss til å handle raskt - det er mye mer effektivt å begynne å forberede seg i dag ved å styrke den digitale kompetansen, implementere moderne løsninger og bygge en datasentrisk kultur.

En av de siste viktige teknologiske hindringene for den storstilte digitale transformasjonen av byggebransjen, som vil påvirke alle selskaper i de kommende årene, er problemet med automatisk klassifisering av elementer i byggeprosjekter.

Uten pålitelig, nøyaktig og skalerbar klassifisering er det umulig å skape grunnlaget for fullverdig analyse, prosessautomatisering og livssyklusstyring ved hjelp av kunstig intelligens og prediktive modeller. Så lenge objektklassifisering fortsatt er avhengig av manuell tolkning utført av erfarte spesialister - formenn, designere, kalkulatører - har byggebransjen fortsatt et mulighetsvindu. Denne tiden kan brukes til å forberede seg på de uunngåelige endringene: økende krav til åpenhet, demokratisering av verktøy og data, og fremveksten av automatiske klassifiseringssystemer som vil endre spillereglene radikalt.

Automatisk klassifisering av elementer i konstruksjonsverdenen kan i sin kompleksitet sammenlignes med objektgjenkjenning i ubemannede kjøresystemer, som er en av hovedutfordringene. La oss tenke oss en ubemannet bil som kjører fra punkt A til punkt B (fig. 10.1-5). Dagens automatiske kjøresystemer sliter med å klassifisere objekter som gjenkjennes av lidarer og kameraer. Det er ikke nok at bilen bare "ser" en hindring eller et landemerke, den må også kunne gjenkjenne om det er en fotgjenger, et veiskilt eller en søppelbøtte.

Hele byggebransjen står overfor en lignende grunnleggende utfordring. Prosjektelementer - som vinduer, dører eller søyler - kan registreres i dokumentasjon, representeres i CAD-modeller, fotograferes på byggeplassen eller gjenkjennes i punktskyer fra laserskanning. Visuell eller grov geometrisk gjenkjenning er imidlertid ikke nok til å bygge et virkelig automatisert prosjektstyringssystem. Det er nødvendig å sørge for at hvert element kategoriseres nøyaktig og konsekvent i en type som er entydig identifiserbar i alle etterfølgende prosesser - fra kalkyler og spesifikasjoner til logistikk, lagerstyring og, viktigst av alt, drift (fig. 4.2-6).

Det er på dette stadiet - i overgangen fra gjenkjenning til meningsfull klassifisering - at en av de største hindringene oppstår. Selv om digitale systemer teknisk sett er i stand til å skille ut og identifisere objekter i modeller og på byggeplassen, ligger hovedproblemet i en korrekt og konteksttuelt stabil definisjon av elementtypen for ulike programvaremiljøer.. For eksempel kan en dør være merket av konstruktøren i en CAD-modell som et element i kategorien "dør", men når den overføres til et ERP- eller PMIS-system, kan den være feilaktig merket, enten på grunn av en feil fra konstruktørens side eller på grunn av uoverensstemmelser mellom systemene. Dessuten mister elementet ofte noen viktige attributter eller forsvinner fra systemregnskapet under dataeksport og -import. Dette fører til et gap i dataflyten og undergraver prinsippet om ende-til-ende-digitalisering av byggeprosesser. Dette skaper et kritisk gap mellom "synlig" og "forståelig" semantisk mening, noe som undergraver dataintegriteten og i betydelig grad kompliserer automatiseringen av prosesser gjennom hele livssyklusen til et byggeprosjekt.

Løsningen på utfordringen med universell klassifisering av bygningselementer ved hjelp av stordata og maskinlæringsteknologi (fig. 10.1-6) vil være en katalysator for transformasjon i hele bransjen - og kanskje en uventet oppdagelse for mange byggefirmaer. Et enhetlig, lærbart klassifiseringssystem vil danne grunnlaget for skalerbar analyse, digital styring og bruk av kunstig intelligens i den daglige driften av byggevirksomheter.

NVIDIA og andre teknologiledere leverer allerede løsninger i andre bransjer som automatisk kan kategorisere og strukturere store mengder tekstlig og visuell informasjon.

NVIDIA's NeMo Curator-modell [161] spesialiserer seg for eksempel på automatisk klassifisering og kategorisering av data i forhåndsdefinerte kategorier, noe som spiller en nøkkelrolle i optimaliseringen av informasjonsbehandlingsrørledninger for finjustering og forhåndstrenings generative AI-modeller. Cosmos-plattformen er trent på virkelige video- og 3D-scener [162], og danner grunnlaget for autonome systemer og digitale tvillinger som allerede er under utvikling i NVIDIA's økosystem. NVIDIA Omniverse, som innen 2025 har blitt det ledende verktøyet for arbeid med USD-formatet, en universell scenebeskrivelse som på sikt kan erstatter IFC-formatet i prosesser for overføring av designinformasjon. Sammen med Isaac Sim, en simulator for robotprosesser [163], representerer løsninger som NeMo Curator, Cosmos og Omniverse et nytt nivå av automatisering: fra datarensing og -filtrering til generering av treningssett, modellering av objektegenskaper og robotopplæring på byggeplassen. Og alle disse verktøyene er gratis og har åpen kildekode, noe som reduserer hindringene for å ta dem i bruk i ingeniør- og bygg- og anleggsbransjen betydelig.

Automatisk klassifisering av data på nivå med strukturerte tabeller er ikke så vanskelig som det kan virke ved første øyekast. Som vi viste i forrige kapittel (fig. 9.1-10), er det mulig å kompensere for manglende eller feilaktige klasseverdier på grunnlag av lignende parametere for andre elementer hvis det finnes akkumulerte historiske data. Hvis elementer med lignende egenskaper allerede har blitt klassifisert riktig i flere fullførte prosjekter, kan systemet med stor sannsynlighet foreslå en passende verdi for et nytt eller ufullstendig element (fig. 10.1-6). En slik logikk, basert på gjennomsnittsverdier og kontekstanalyse, kan være spesielt effektiv ved massebehandling av tabelldata som kommer fra estimater, spesifikasjoner eller CAD-modeller.

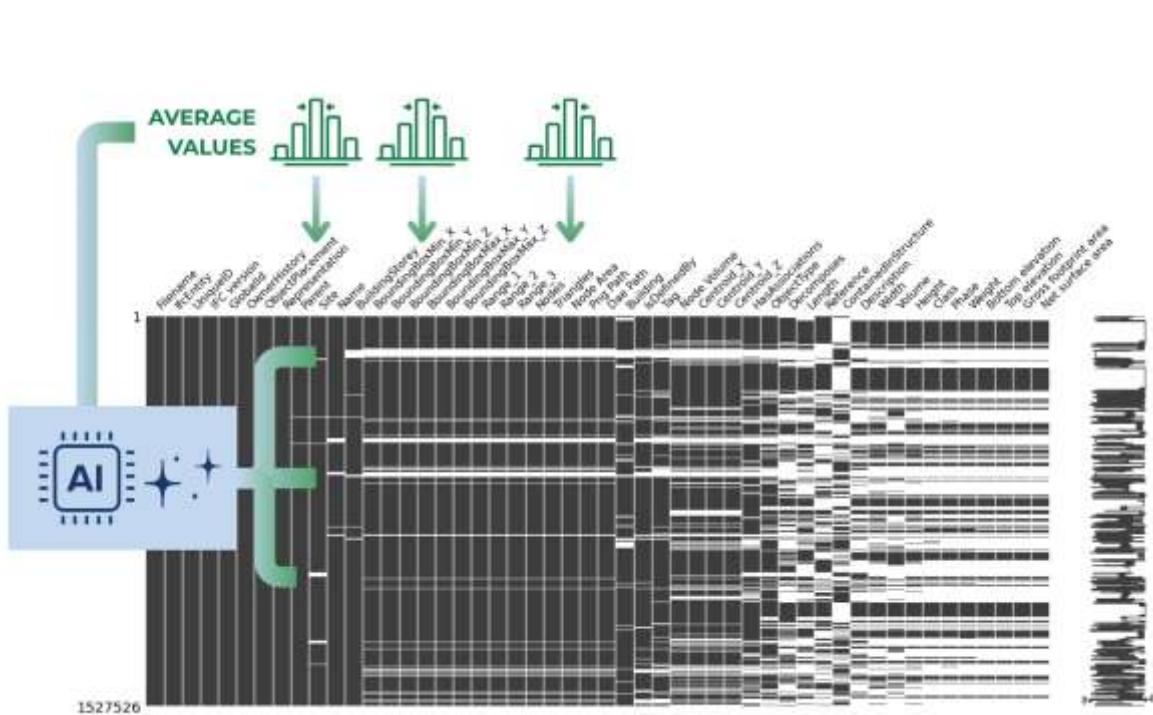
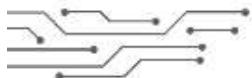


Fig. 10.1-6 Maskinlæring kan hjelpe deg med å automatisk finne gjennomsnittsverdier for utfylte (hvite bokser) tabellparametere basert på tidligere prosjekter.

På bakgrunn av den raske utviklingen innen maskinlæring er det klart: I 2025 er det naivt å tro at problemet med automatisk klassifisering av bygningselementer vil forbli uløst i lang tid fremover. Ja, dagens algoritmer har ennå ikke nådd full modenhet, spesielt med ufullstendige eller heterogene data, men muligheten for tilpasning er i ferd med å lukke seg.

Selskaper som allerede har investert i innsamling,rensing og organisering av data, og som tar i bruk ETL-automatiseringsverktøy, vil ha et fortrinn. Resten risikerer å sakke akterut - akkurat som selskaper som en gang i tiden ikke klarte å møte utfordringene med digital transformasjon i transport- og finansbransjen

De som fortsetter å basere seg på manuell datahåndtering og tradisjonelle metoder for kostnads- og tidsestimering, risikerer å havne i samme situasjon som drosjeflåtene på 2000-tallet, ute av stand til å tilpasse seg en tid med mobilapper og automatiserte ruteberegninger på begynnelsen av 2020-tallet.



KAPITTEL 10.2.

EN PRAKTISK GUIDE TIL IMPLEMENTERING AV EN DATADREVET TILNÆRMING

Fra teori til praksis: Et veikart for digital transformasjon i byggebransjen

Byggebransjen er gradvis på vei inn i en ny utviklingsfase, der kjente prosesser i økende grad suppleres - og noen ganger til og med erstattes - av digitale plattformer og transparente samhandlingsmodeller. Dette bryr ikke bare på utfordringer, men også betydelige muligheter. De organisasjonene som allerede i dag bygger opp en langsigkt digital strategi, vil ikke bare kunne opprettholde sin posisjon i markedet, men også utvide den ved å tilby kundene moderne tilnærmlinger og pålitelige, teknologisk støttede løsninger.

Det er viktig å innse at kunnskap om konsepter og teknologi bare er et utgangspunkt. Ledere og spesialister står overfor et praktisk spørsmål: Hvor skal man starte implementeringen, og hvordan skal man gjøre teoretiske ideer om til reell verdi? I tillegg melder spørsmålet seg i økende grad: Hva skal virksomheten baseres på, hvis tradisjonelle metoder for kostnadsberegning og tidsplanlegging kan revideres av kunden når som helst.

Svaret ligger sannsynligvis ikke så mye i teknologien, men i dannelsen av en ny yrkeskultur der arbeid med data oppfattes som en integrert del av den daglige praksisen. Det er den manglende oppmerksomheten rundt digital teknologi og innovasjon som har ført byggebransjen inn i den alvorlige tilbakeliggelsen som har blitt observert de siste tiårene [43].

Ifølge McKinsey utgjør FoU-utgiftene i byggebransjen mindre enn 1% av omsetningen, mens de i bil- og flyindustrien ligger på 3,5-4,5%. På samme måte ligger IT-kostnadene i byggebransjen på under 1% av de totale inntektene [107].

Resultatet er at ikke bare automatiseringsgraden, men også arbeidsproduktiviteten i byggebransjen synker, og i 2020 vil bygningsarbeideren produsere mindre enn for et halvt århundre siden (fig. 10.2-1)

Slike produktivitetsproblemer i bygg- og anleggssektoren er felles for de fleste industrieland og utviklingsland (produktiviteten i bygg- og anleggssektoren har falt i 16 av 29 OECD-land (figur 2.2-1)), og de peker ikke bare på mangel på teknologi, men også på behovet for systemiske endringer i selve tilnærmingen til ledelse, opplæring og innovasjon.

Hvor vellykket den digitale transformasjonen blir, avhenger ikke så mye av antall og tilgjengelighet av verktøy, men av organisasjonenes evne til å gjennomgå prosessene sine og utvikle en kultur som er åpen for endring. Det er ikke teknologien i seg selv som er nøkkelen, men menneskene og prosessene som sikrer effektiv bruk, støtter kontinuerlig læring og oppmuntrer til aksept av nye ideer.

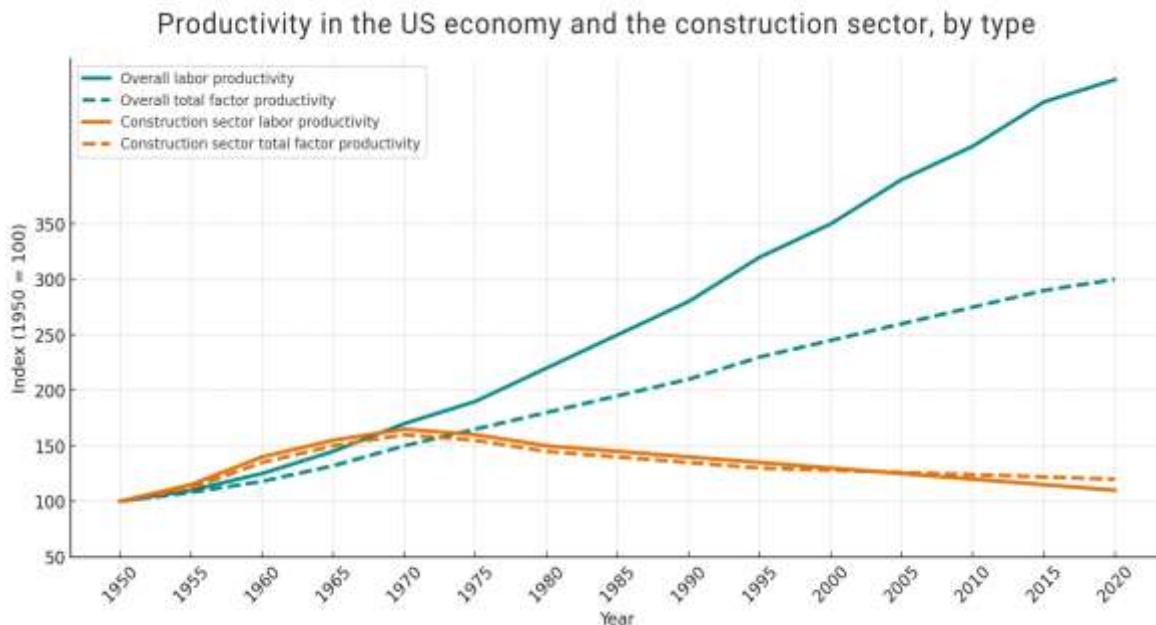


Fig. 10.2-1 Paradokset med arbeidsproduktivitet og total ressursproduktivitet i den amerikanske økonomien og bygg- og anleggssektoren (1950-2020) (basert på [43]).

I de første delene av boken ble forretningsmiljømodellen sammenlignet med et skogøkosystem (figur 2.1-2, figur 1.2-4, figur 1.3-2). I en sunn skog spiller periodiske branner, til tross for all sin ødeleggende kraft, en nøkkelrolle i den langsiktige fornyelsen. De rydder jorden for gammel vegetasjon, gir tilbake lagrede næringsstoffer og skaper rom for nytt liv. Noen plantearter har til og med utviklet seg slik at frøene deres bare åpner seg når de utsettes for høye branntemperaturer - en naturlig mekanisme som gir den ideelle tiden for spiring.

På samme måte kan kriser i næringslivet fungere som en "kontrollert utbrenhet", som oppmuntrer til nye tilnæringer og selskaper som ikke er bundet til utdaterte systemer. Slike perioder tvinger frem en avvikling av ineffektiv praksis, noe som frigjør ressurser til innovasjon. På samme måte som en skog etter en brann starter med pionerplanter, kan en virksomhet etter en krise danne nye, fleksible prosesser som blir grunnlaget for et modent informasjonsmiljø.

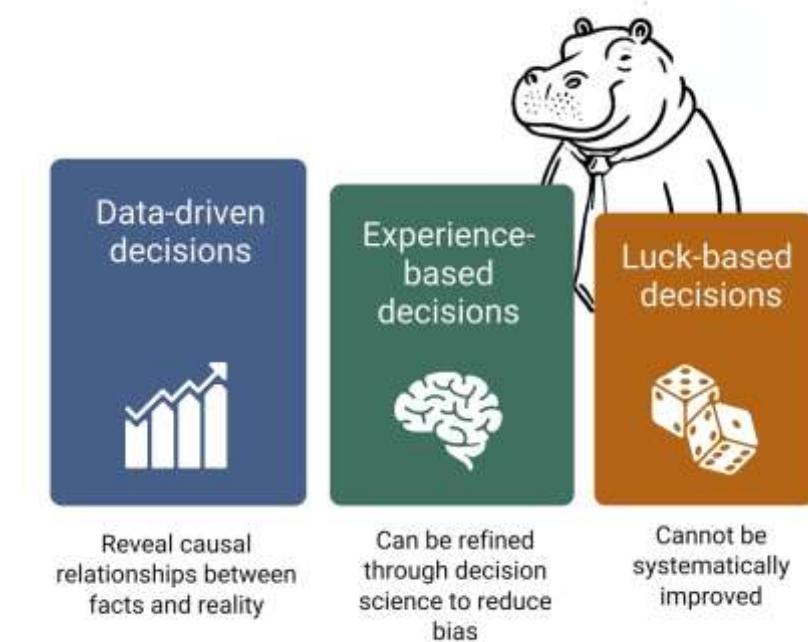
Bedrifter som klarer å tolke disse "signalene" riktig og omdanne den forstyrrende energien til konstruktive endringer, vil nå et nytt prestasjonsnivå - med mer transparente, adaptive dataprosesser som styrker organisasjonens naturlige evne til å fornye seg og vokse.

Det er ikke lenger noen tvil om at kunstig intelligens og maskinlæring får stadig større innflytelse på næringslivet. Dette er ikke bare en midlertidig trend, men en strategisk nødvendighet. Bedrifter som ignorerer kunstig intelligens, risikerer å miste konkurransesevnen i et marked som i stadig større grad oppfordrer til innovasjon og fleksibilitet.

Fremtiden tilhører dem som ikke bare ser på AI som et verktøy, men som en mulighet til å tenke nytt om alle aspekter av virksomheten - fra optimalisering av prosesser til ledelsesbeslutninger.

Å legge det digitale grunnlaget: 1-5 trinn til digital modenhet

I dette kapittelet ser vi på veikartet for digital transformasjon og identifiserer de viktigste stegene som kreves for å implementere en datadrevet tilnærming som kan bidra til å transformere både bedriftskulturen og bedriftens informasjonsøkosystem.



Figur 10.2-2 Kontrollert oppdatering og valg av strategi: case, erfaring eller data.

Ifølge McKinseys studie "Why Digital Strategies Fail" (2018) er det minst fem grunner [164] til at bedrifter ikke lykkes med å nå sine mål for digital transformasjon

- **Uklare definisjoner:** Ledere og ledere har ulike oppfatninger av hva "digital" betyr, noe som fører til misforståelser og uoverensstemmelser.
- **Misforståelse av den digitale økonomien:** Mange selskaper undervurderer omfanget av endringene som digitaliseringen fører med seg for forretningsmodeller og bransjedynamikk (figur 10.1-6).
- **Ignorerer økosystemer:** Bedrifter fokuserer på individuelle teknologiløsninger (datasiloer) og overser behovet for å integrere seg i bredere digitale økosystemer (figur 2.2-2, figur 4.1-12).
- **Undervurdering av konkurrentenes digitalisering:** Ledere tar ikke hensyn til at konkurrentene også aktivt tar i bruk digital teknologi, noe som kan føre til tap av konkurransefortrinn.
- **Digitaliseringens tosidighet:** Administrerende direktører delegerer ansvaret for digital transformasjon til andre ledere, noe som byråkratiserer kontrollen og bremser endringsprosessen.

For å håndtere disse utfordringene kreves det en klar forståelse og tilpasning av digitale strategier på alle nivåer i organisasjonen. Før man bygger en digital strategi, er det viktig å forstå utgangspunktet. Mange organisasjoner har en tendens til å ta i bruk nye verktøy og plattformer uten å ha et fullstendig bilde av dagens situasjon.

Trinn 1: Gjennomfør en revisjon av de nåværende systemene og dataene dine.

Før du endrer prosesser, er det viktig å forstå hva som allerede er på plass. Ved å gjennomføre en revisjon kan du identifisere svakheter i datahåndteringen og forstå hvilke ressurser som kan brukes. En revisjon er et slags "røntgenbilde" av forretningsprosessene dine. Den gjør det mulig å identifisere risikoområder og avgjøre hvilke data som er kritiske for prosjektet eller virksomheten, og hvilke som er sekundære.

Hovedhandlinger:

- Kartlegg IT-miljøet ditt (i Draw.io, Lucidchart, Miro, Visio eller Canva). List opp systemene som brukes (ERP, CAD, CAFM, CPM, SCM og andre) som er involvert i prosessene dine, og som vi diskuterte i kapittelet "Teknologier og styringssystemer i moderne bygg og anlegg" (fig. 1.2-4)
- Vurder datakvalitetsproblemer for hvert system med tanke på hyppigheten av duplikater, mulige manglende verdier og formatinkonsistens i hvert system.
- Identifiser "smertepunkter" - steder der prosesser kan bryte sammen eller ofte krever manuell inngrip - import, eksport og ytterligere valideringsprosesser.

Hvis du vil at teamet skal stole på rapportene, må du sørge for at dataene er korrekte fra starten av.

En kvalitetsdatarevisjon vil vise deg hvilke data:

- Trenger videreutvikling (automatiske rengjøringsprosesser eller ytterligere transformasjoner må settes opp)
- De er "søppel" som bare tetter igjen systemene, og som man kan kvitte seg med ved å ikke bruke dem i prosessene lenger.

Det er mulig å gjennomføre en slik revisjon på egen hånd. Men noen ganger er det nyttig å engasjere en ekstern konsulent - spesielt fra andre bransjer: Et nytt perspektiv og uavhengighet fra "særegenheter" i byggebransjen vil bidra til å vurdere status quo på en nøktern måte og unngå de typiske fallgruvene ved å være forutinntatt til fordel for bestemte løsninger og teknologier.

Trinn 2: Identifiser viktige standarder for dataharmonisering.

Etter revisjonen er det nødvendig å lage felles regler for hvordan man arbeider med data. Som vi diskuterte i kapittelet "Standarder: Fra tilfeldige filer til en veldesignet datamodell", vil dette bidra til å eliminere siloflyt av data.

Uten en felles standard vil hvert team fortsette å jobbe på "sin egen måte", og du vil opprettholde en "dyrehage" av integrasjoner der data går tapt ved hver eneste konvertering.

Hovedhandlinger:

- Velg datastandarder for utveksling av informasjon mellom systemer:
 - For tabelldata kan dette være strukturerte formater som CSV, XLSX eller mer effektive

formater som Parquet

- For utveksling av løst strukturerte data og dokumenter: JSON eller XML

■ Beherske arbeidet med datamodeller:

- Begynn med å parametrisere oppgavene på nivå med den konseptuelle datamodellen - som beskrevet i kapittelet "Datamodellering: konseptuell, logisk og fysisk modell" (Fig. 4.3-2)
- Etter hvert som du går dypere inn i forretningsprosessens logikk, går du over til å formalisere kravene ved hjelp av parameter i de logiske og fysiske modellene (fig. 4.3-6)
- Identifisere viktige enheter, deres attributter og relasjoner i prosesser, og visualisere disse relasjonene - både mellom enheter og mellom parameter (figur 4.3-7)

■ Bruk regulære uttrykk (RegEx) for å validere og standardisere data (figur 4.4-7), som vi diskuterte i kapittelet "Strukturerte krav og regulære RegEx-uttrykk". RegEx er ikke et komplekst, men et svært viktig tema i arbeidet med å lage krav på nivå med fysiske datamodeller.

Uten standarder på datanivå og prosessvisualisering er det umulig å levere et konsistent og skalerbart digitalt miljø. Husk at "dårlige data er dyre". Og kostnadene ved feil øker etter hvert som et prosjekt eller en organisasjon blir mer kompleks. Å forene formater og definere regler for navngivning, struktur og validering er en investering i fremtidige løsningers stabilitet og skalerbarhet.

Trinn 3. Implementer DataOps og automatiser prosesser.

Uten en veldefinert arkitektur vil bedrifter uunngåelig stå overfor uensartede data i siloformede informasjonssystemer. Dataene vil være uintegrerte, duplisert på flere steder og kostbare å vedlikeholde.

Forestill deg at data er vann, og at dataarkitekturen er det komplekse systemet av rørledninger som transporterer vannet fra lagringskilden til der det skal brukes. Det er dataarkitekturen som bestemmer hvordan informasjon samles inn, lagres, omdannes, analyseres og leveres til sluttbrukere eller applikasjoner.

DataOps (Data Operations) er en metodikk som integrerer innsamling, rensing, validering og bruk av data i én enkelt automatisert prosessflyt, slik vi diskuterte i detalj i del 8 av boken.

Hovedhandlinger:

■ Opprett og tilpass ETL -transportører for å automatisere prosesser:

- Extract: Organiser automatisk datainnsamling fra PDF -dokumenter (fig. 4.1-2, fig. 4.1-5, fig. 4.1-7), Excel -regneark, CAD -modeller (fig. 7.2-4), ERP -systemer og andre kilder du jobber med.
- Transform: Sett opp automatiske prosesser for å transformere data til et enkelt strukturert format og automatisere beregninger som vil finne sted utenfor lukkede applikasjoner (Fig. 7.2-8)
- Last inn: prøv å opprette en automatisk opplasting av data til oversiktstabeller, dokumenter eller sentraliserte arkiver (fig. 7.2-9, fig. 7.2-13, fig. 7.2-16)

- Automatiser kalkulasjons- og QTO-prosessen (Quantity Take-Off), som vi beskrev i kapittelet "QTO Quantity Take-Off: Gruppering av prosjektdata etter attributter":
 - Sett opp automatisk utvinning av volumer fra CAD -modeller ved hjelp av API-er, plugin-moduler eller verktøy for omvendt konstruksjon (fig. 5.2-5).
 - Lag regler for gruppering av elementer for ulike klasser etter attributter i form av tabeller (figur 5.2-12)
 - Forsøk å automatisere ofte gjentatte volum- og kostnadsberegninger utenfor modulære, lukkede systemer (fig. 5.2-15)
- Begynn å bruke Python og Pandas til å behandle data, som vi diskuterte i kapittelet "Python Pandas: et uunnværlig verktøy for arbeid med data":
 - Bruk DataFrame til å arbeide med XLSX-filer og automatisere behandlingen av tabelldata (Fig. 3.4-6)
 - Automatiser aggregering og transformasjon av informasjon ved hjelp av ulike Python-biblioteker
 - Bruk LLM for å gjøre det enklere å skrive ferdige kodeblokker og hele Pipelines (figur 7.2-18)
 - Prøv å lage en Pipeline i Python, som finner feil eller ser avvik og sender et varsel til den ansvarlige personen (f.eks. prosjektlederen) (figur 7.4-2)

Automatisering basert på DataOps-prinsipper gjør at du kan gå fra manuell og fragmentert datahåndtering til bærekraftige og repeterbare prosesser. Dette reduserer ikke bare belastningen på de ansatte som håndterer de samme transformasjonene hver dag, men øker også påliteligheten, skalerbarheten og gjennomsiktigheten til hele informasjonssystemet dramatisk.

Trinn 4: Opprett et økosystem for åpen datastyring.

Til tross for utviklingen av lukkede modulære systemer og integrasjonen av dem med nye verktøy, står bedriftene overfor et alvorlig problem - den økende kompleksiteten i slike systemer overgår nytten av dem. Den opprinnelige ideen om å skape én proprietær plattform som dekker alle forretningsprosesser, har ført til en overdreven sentralisering, der enhver endring krever betydelige ressurser og tid for å tilpasse seg.

Som vi diskuterte i kapittelet "Mycelium i bedriften: Hvordan data kobler sammen forretningsprosesser", krever effektiv datahåndtering et åpent og enhetlig økosystem som knytter sammen alle informasjonskilder.

Nøkkelementer i økosystemet:

- Velg et passende datalager:
 - For tabeller og beregninger bruker du databaser - for eksempel PostgreSQL eller MySQL (Fig. 3.1-7)
 - For dokumenter og rapporter kan skylagring (Google Drive, OneDrive) eller systemer som støtter JSON-format være egnet
 - Sjekk ut mulighetene som ligger i datavarehus, datasjøer og andre verktøy for sentralisert lagring og analyse av store mengder informasjon (figur 8.1-8)

■ Implementere løsninger for å få tilgang til proprietære data:

- Hvis du bruker proprietære systemer, må du konfigurere tilgang til dem via API eller SDK for å hente data for ekstern behandling (fig. 4.1-2)
- Gjør deg kjent med potensialet i reverse engineering-verktøy for CAD formater (figur 4.1-13)
- Sett opp ETL-rørledninger som med jevne mellomrom samler inn data fra applikasjoner eller servere, konverterer dem til åpne strukturerte formater og lagrer dem i repositorier (figur 7.2-3)
- Diskutere i teamet hvordan man kan gi tilgang til data uten behov for proprietær programvare
- Husk: Data er viktigere enn grensesnitt. Det er strukturen og tilgjengeligheten av informasjon, ikke spesifikke brukergrensesnittverktøy, som gir langsiktig verdi

■ Tenk på å opprette et kompetansesenter (CoE) for data, slik vi diskuterte i kapittelet "Kompetansesenter (CoE) for datamodellering", eller hvordan du kan tilby datakompetanse på andre måter (figur 4.3-9)

Økosystemet for datahåndtering skaper et enhetlig informasjonsrom der alle prosjektdeltakere arbeider med konsistent, oppdatert og verifisert informasjon. Det er grunnlaget for skalarbare, fleksible og pålitelige digitale prosesser.

Frigjør potensialet i data: 5-10 trinn til digital modenhet

I tillegg til teknisk integrasjon er sluttbrukernes aksept av digitale løsninger en viktig faktor for å lykkes med implementeringen. Å engasjere kunder eller brukere i resultatmåling er både en utfordring når det gjelder å forbedre brukeropplevelsen og å håndtere endringer i bedriften. Hvis en løsning ikke passer inn i en kjent arbeidsflyt eller ikke løser reelle bruker- eller kundeproblemer, vil den ikke bli brukt, og det kan ingen ekstra tiltak og incentiver gjøre noe med.

Transformasjon er en iterativ prosess basert på analyse av data om brukerinteraksjon med nye prosesser, med hyppige testsyklinger, stadige tilbakemeldinger og forbedringer.

Trinn 5: Bygg en datakultur, lær opp personalet og innhent tilbakemeldinger

Selv det mest avanserte systemet vil ikke fungere uten de ansattes engasjement. Du må skape et miljø der dataene brukes daglig, og der teamet forstår verdien av dem.

I den britiske regjeringens rapport "Data Analytics and AI in Government Project Delivery", som ble publisert i 2024, bemerkes det [83] at opplæring av fagfolk med nødvendig kompetanse innen databehandling og tolkning er avgjørende for en vellykket implementering av dataanalyse og AI.

Mangel på dataanalysekjennskap er et av de viktigste hindrene for digital transformasjon. Ledere er vant til etablerte rutiner: kvartalsvise sykluser, prioriterte initiativer og tradisjonelle måter å drive prosjekter fremover på. Endring krever en spesiell leder - høyt nok oppe i hierarkiet til å ha innflytelse, men ikke så høyt oppe at han eller hun har tid og motivasjon til å lede et langsiktig transformasjonsprosjekt.

Hovedhandlinger:

- Erkjennelsen av at det er nødvendig å gå fra subjektive beslutninger basert på en høyt lønnet ansatts mening (HiPPO) til en beslutningskultur basert på fakta og data, som omtalt i kapittelet "HiPPO or the Danger of Opinion in Decision Making" (figur 2.1-9).
- Organiser systematisk opplæring:
 - Hold opplæring i bruk av strukturerte data, og inviter eksperter fra andre bransjer som ikke har fordommer mot produkter og konsepter som er populære i byggebransjen i dag
 - Diskutere dataanalysetilnæringer og -verktøy med kolleger, og selvstendig lære praktisk arbeid med verktøy som Python, pandas og LLM (Fig. 4.1-3, Fig. 4.1-6)
 - Opprett et bibliotek med veiledninger (gjerne med korte videoer) om strukturering av data (fig. 3.2-15) og oppretting av datamodeller (fig. 4.3-6, fig. 4.3-7)
- Bruk moderne læringsteknologi:
 - Bruk språkmodeller (LLM) til å støtte kode- og datamanipulering, inkludert generering, refaktorisering og analyse av kode, samt behandling og tolkning av tabellinformasjon (figur 3.4-1)
 - Utforsk hvordan LLM-generert kode kan tilpasses og integreres i en komplett Pipeline-løsning når du arbeider i et offline utviklingsmiljø (IDE) (Fig. 4.4-14, Fig. 5.2-13)

Når en leder fortsetter å ta beslutninger på "gamlemåten", vil ingen opplæring overbevise folk om å ta analyser på alvor.

Det er umulig å bygge en datakultur uten kontinuerlig tilbakemelding. Tilbakemeldinger gjør det mulig å avdekke mangler i prosesser, verktøy og strategier som ikke kan oppdages gjennom interne rapporter eller formelle KPI-er metrics. Komplimenterende kommentarer fra brukere av løsningene dine vil ikke gi praktisk verdi. Det er kritiske tilbakemeldinger som er verdifulle, særlig hvis de er basert på konkrete observasjoner og fakta. Men det krever innsats å innhente slik informasjon: Du må bygge opp prosesser der deltakerne - både interne og eksterne - kan dele kommentarer (det kan være fornuftig å gjøre det anonymt) uten forvrengning og uten frykt for at meningene deres kan påvirke deres eget arbeid. Det er viktig at de gjør det uten forvrengning og uten frykt for negative konsekvenser for dem selv.

All læring er til syvende og sist selvlæring [165].

- Milton Friedman, amerikansk økonom og statistiker

Implementering av analyseverktøy bør ledsages av regelmessig verifisering av effektiviteten i praksis (ROI, KPI-er), noe som bare kan oppnås gjennom struktureret tilbakemelding fra ansatte, kunder og partnere. På denne måten kan bedriften ikke bare unngå å gjenta feil, men også tilpasse seg raskere til endringer i omgivelsene. Å ha en mekanisme for innsamling og analyse av tilbakemeldinger er et av tegnene på modenhet i en organisasjon som beveger seg fra sporadiske digitale initiativer til en bærekraftig modell for kontinuerlig forbedring (figur 2.2-5).

Trinn 6: Fra pilotprosjekter til oppskalering

Velg kamper som er store nok til å bety noe, og små nok til å vinne.

- Jonathan Kozol

Det er ekstremt risikabelt å sette i gang digital transformasjon "på en gang og overalt". En mer effektiv tilnærming er å starte med pilotprosjekter og gradvis skalere opp vellykkede metoder.

Hovedhandlinger:

■ Velg riktig prosjekt for piloten:

- Definer et spesifikt forretningsmål eller en spesifikk prosess med målbare resultater (KPI, ROI) (Fig. 7.1-5)
- Velg en ETL-automatiseringsprosess, for eksempel automatisk datavalidering eller beregning av arbeidsmengde (QTO hjelpe med Python og Pandas (figur 5.2-10))
- Fastsette klare mål for suksess (f.eks. redusere tiden det tar å utarbeide inspeksjonsspesifikasjoner eller datavalidiseringsrapporter fra en uke til en dag)

■ Bruk iterative tilnærninger:

- Begynn med enkle datakonverteringsprosesser, og lag strømmekonverteringer av data i flere formater til de formatene du trenger for prosessene dine (fig. 4.1-2, fig. 4.1-5)
- Øk gradvis kompleksiteten i oppgavene og utvid automatiseringen av prosessene ved å danne en komplett pipeline i IDE basert på dokumenterte kodeblokker (fig. 4.1-7, fig. 7.2-18).
- Dokumenter og registrer (gjerne med korte videoer) vellykkede løsninger, og del dem med kolleger eller i fagmiljøer

■ Utvikle maler og tilhørende dokumentasjon for å replikere slike løsninger, slik at de kan brukes effektivt av dine kolleger (eller medlemmer av fagmiljøet, inkludert brukere av sosiale medier)

En trinnvis "opprulling" gjør det mulig å opprettholde den høye kvaliteten på endringene og ikke falle inn i kaoset med parallelle implementeringer. "Fra liten til stor"-strategien minimerer risikoen og gjør at du kan lære av små feil uten at de vokser til kritiske problemer.

Ved å gå fra en prosjektilnærming, der medarbeiderne bare er delvis involvert, til å danne permanente team (f.eks. kompetansesentre - CoE-er) kan man sikre en bærekraftig utvikling av produktet også etter at den første versjonen er lansert. Slike team støtter ikke bare eksisterende løsninger, men fortsetter også å forbedre dem.

Dette reduserer avhengigheten av langvarige godkjenninger: Teammedlemmene får fullmakt til å ta beslutninger innenfor sitt eget ansvarsområde. Resultatet er at lederne slipper å behovet for å detaljstyre,

og teamene kan fokusere på å skape reell verdi.

Å utvikle nye løsninger er ikke en sprint, men et maratonløp. De som lykkes med det, er de som i utgangspunktet fokuserer på langsiktig og konsekvent arbeid.

Det er viktig å være klar over at teknologi krever konstant utvikling. Å investere i langsiktig utvikling av teknologiske løsninger er grunnlaget for en vellykket virksomhet.

Trinn 7: Bruk åpne dataformater og -løsninger

Som vi diskuterte i kapitlene om modulære plattformer (ERP, PMIS, CAFM, CDE osv.), er det viktig å fokusere på åpne og universelle dataformater som sikrer uavhengighet fra leverandør løsninger og øker tilgjengeligheten av informasjon for alle interesserter.

Hovedhandlinger:

- Gå fra lukkede formater til åpne formater:
 - Bruk åpne formater i stedet for proprietære formater, eller finn en måte å sette opp automatisk opplasting eller konvertering av lukkede formater til åpne formater (figur 3.2-15).
 - Implementere verktøy for å arbeide med Parquet, CSV, JSON, XLSX, som er utvekslingsstandardene mellom de fleste moderne systemer (fig. 8.1-2)
 - Hvis arbeid med 3D-geometri spiller en viktig rolle i prosessene dine, bør du vurdere å bruke åpne formater som USD, glTF, DAE eller OBJ (figur 3.1-14).
- Bruk vektordatabaser databaser for effektiv analyse og gjenfinning av informasjon:
 - Bruk Bounding Box og andre metoder for å forenkle 3D -geometrien (figur 8.2-1)
 - Tenk over hvor du kan implementere datavektorisering - konvertere tekster, objekter eller dokumenter til numeriske representasjoner (figur 8.2-2)
- Bruk verktøy for stordataanalyse:
 - Organiser lagring av akkumulerte historiske data (f.eks. PDF, XLSX, CAD) i formater som egner seg for analyse (Apache Parquet, CSV, ORC) (fig. 8.1-2).
 - Begynne å anvende grunnleggende statistiske metoder og arbeide med representative utvalg - eller, som et minimum, gjøre deg kjent med de grunnleggende prinsippene for statistikk (figur 9.2-5)
 - Implementer og lær deg verktøy for å visualisere data og forholdet mellom data for å synliggjøre resultatene av analysen. Uten god visualisering er det umulig å forstå selve dataene eller prosessene som er basert på dem (figur 7.1-4).

Overgangen til åpne dataformater og innføring av verktøy for analyse, lagring og visualisering av informasjon legger grunnlaget for en bærekraftig og uavhengig digital forvaltning. Dette reduserer ikke bare avhengigheten av leverandører, men sikrer også lik tilgang til data for alle interesserter.

Trinn 8. Begynn å implementere maskinlæring for prediksjon

Mange bedrifter har samlet opp enorme mengder data - en slags "informasjonsgeysirer" som fortsatt står ubrukt. Disse dataene har blitt samlet inn i hundrevis eller tusenvis av prosjekter, men har ofte bare blitt brukt én gang eller ikke i det hele tatt i videre prosesser. Dokumenter og modeller som er lagret i lukkede formater og systemer, blir ofte oppfattet som foreldet og ubruklig. I virkeligheten er de imidlertid en verdifull ressurs - grunnlaget for å analysere feil som er gjort, automatisere rutineoperasjoner og utvikle innovative løsninger for automatisk klassifisering og funksjonsgjenkjenning i fremtidige prosjekter.

Hovedutfordringen er å lære seg hvordan man kan hente ut disse dataene og omdanne dem til nyttig innsikt. Som omtalt i kapittelet om maskinlæring og prediksjoner, har maskinlæringsteknikker potensial til å forbedre nøyaktigheten av estimatorer og prediksjoner i en rekke ulike byggrelaterte prosesser. Ved å utnytte de akkumulerte dataene fullt ut kan man forbedre effektiviteten, redusere risikoen og bygge opp bærekraftige digitale prosesser.

Hovedhandlinger:

- Begynn med enkle algoritmer:
 - Prøv å bruke lineær regresjon - ved hjelp av tips fra LLM - for å forutsi tilbakevendende ytelse i datasett der avhengigheten av et stort antall faktorer er fraværende eller minimal (figur 9.3-4)
 - Tenk over i hvilke faser av prosessene dine k-nærmeste naboer-algoritmen (k-NN) teoretisk sett kan brukes - for eksempel til klassifiseringsoppgaver, vurdering av objektlikhet eller prediksjon basert på historiske analogier (fig. 9.3-5).
- Samle inn og strukturere data for å trenere modeller:
 - Samle historiske prosjektdata på ett sted og i ett enkelt format (figur 9.1-10)
 - Arbeid med kvaliteten og representativiteten til treningsseksempler, gjennom automatisert ETLfigur 9.2-8)
 - Lær deg å dele dataene inn i trenings- og testsett, slik vi gjorde i eksemplet med Titanic-datasettet (figur 9.2-6, figur 9.2-7)
- Vurdere å utvide bruken av maskinlæringsteknikker til et bredt spekter av oppgaver, fra å forutsi tidslinjer for prosjekter til å optimalisere logistikk, ressursstyring og tidlig identifisering av potensielle problemer

Maskinlæring er et verktøy som gjør arkiverte data til en verdifull ressurs for prediksjon, optimalisering og informert beslutningstaking. Begynn med små datasett (figur 9.2-5) og enkle modeller, og øk gradvis kompleksiteten.

Trinn 9. Integrerer IoT og moderne datainnsamlingsteknologi

Bygg- og anleggsbransjen er i ferd med å digitaliseres: Hvert eneste bilde og hvert eneste innlegg i Teams er allerede en del av en større prosess der virkeligheten parametriseres og tokeniseres. På samme måte som GPS en gang forandret logistikken, er IoT, RFID og automatisk datainnsamlingsteknologi i ferd med å endre byggebransjen. Som vi diskuterte i kapittelet "Tingenes internett og smarte kontrakter", er den digitale byggeplassen med sensorer og automatisert overvåking fremtiden for bransjen.

Hovedhandlinger:

- Implementere IoT -enheter, RFID -tagger og beskrive prosessene knyttet til dem:
 - Evaluér hvilke områder eller prosjektfaser i et prosjekt som kan få størst avkastning på investeringen (ROI) - for eksempel for overvåking av temperatur, vibrasjoner, fuktighet eller
 - Vurder å bruke RFID til å spore materialer, verktøy og utstyr gjennom hele forsyningsskjeden
 - Vurder hvordan de innsamlede dataene kan integreres i et enkelt informasjonssystem, for eksempel Apache NiFi, for automatisert sanntidsbehandling og analyse (figur 7.4-5)
- Etablere et system for sanntidsovervåking:
 - Utvikle dashbord for å spore viktige prosess- eller prosjektmålinger ved hjelp av visualiseringsverktøy som Streamlit, Flask eller Power BI)
 - Sett opp automatiske varsler for å signalisere kritiske avvik fra planen eller normene (fig. 7.4-2)
 - Vurdere potensialet for forebyggende vedlikehold av utstyr basert på innsamlede data og identifiserte mønstre (figur 9.3-6)
- Kombiner data fra ulike kilder:
 - Begynn med å visualisere datamodellen på fysisk nivå - gjenspeil strukturen i informasjonsflyten og nøkkelparametere som kommer fra CAD-systemer, IoT-enheter og ERP-plattformer (fig. 4.3-1)
 - Begynn med å lage en grov skisse av en enhetlig plattform for dataanalyse og beslutningsstøtte for ledelsen. Kartlegg nøkkelfunksjoner, datakilder, brukere og forventede bruksscenarioer (figur 4.3-7).

Jo raskere du begynner å koble reelle prosesser til den digitale verdenen, desto raskere kan du styre dem med data - effektivt, transparent og i sanntid.

Trinn 10. Forbered deg på fremtidige endringer i bransjen

Bygg- og anleggsbedrifter er under konstant press fra omgivelsene: økonomiske kriser, teknologiske sprang og endringer i regelverket. I likhet med en skog som må tåle regn, snø, tørke og stekende sol, lever bedriftene under forhold som krever kontinuerlig tilpasning. Og på samme måte som trær blir motstandsdyktige mot frost og tørke gjennom dype rotssystemer, er det bare de organisasjonene som har et solid fundament av automatiserte prosesser, evnen til å forutse endringer og fleksibelt tilpassede strategier, som forblir levedyktige og konkurransedyktige.

Som nevnt i kapittelet "Overlevelsesstrategier: Å bygge konkurransefortrinn", er byggebransjen på vei inn i en fase med radikale endringer. Samspillet mellom byggherre og entreprenør beveger seg i retning av en modell for overtalelse, der åpenhet, forutsigbarhet og digitale verktøy erstatter tradisjonelle tilnærminger. I denne nye virkeligheten er det ikke de største, men de mest fleksible og teknologisk modne som vinner.

Hovedhandlinger:

- Analyser sårbarheter i virksomheten i sammenheng med åpne data:
 - Vurder hvordan demokratisering av tilgangen til data som en del av uberiseringen kan ha en ødeleggende innvirkning på konkurransesfortrinnet og virksomheten din (figur 10.1-5)
 - Tenk på en strategi for å gå fra ugjennomsiktige og siloformede prosesser til forretningsmodeller basert på åpne løsninger, systeminteroperabilitet og datatransparens (fig. 2.2-5).
- Utvikle en langsiktig digital strategi:
 - Bestem deg for om du ønsker å være en innovasjonsleder eller om du foretrekker et "innhentingsscenario" der du sparar på ressursene dine
 - Beskriv stadiene: kort sikt (prosessautomatisering, datastrukturering), mellomlang sikt (implementering av LLM og ETL), lang sikt (digitale økosystemer, sentraliserte repositorier)
- Tenk på å utvide porteføljen av tjenester:
 - Vurder å tilby nye tjenester (med fokus på energieffektivitet, ESG, datatjenester). Vi skal snakke mer om nye forretningsmodeller i neste kapittel.
 - Forsök å posisjonere deg som en pålitelig teknologipartner som støtter hele livssyklusen til et anlegg, fra design til drift. Tilliten til deg bør være basert på en systematisk tilnærming, transparente prosesser og evnen til å levere bærekraftige teknologiske løsninger.

I et transformasjonsmiljø er det ikke de som bare reagerer på endringer som vinner, men de som er proaktive. Fleksibilitet, åpenhet og digital modenhet er grunnlaget for å bygge morgendagen på en bærekraftig måte.

Veikart for transformasjon: fra kaos til datadrevet selskap

Følgende plan kan fungere som et første referansepunkt - et utgangspunkt for å utforme din egen datadrevne strategi for digital transformasjon:

- **Revisjon og standarder:** analysere nåværende tilstand, samkjøre data
- **Datastrukturering og -klassifisering:** automatiser omformingen av ustukturerte og løst strukturerte data
- **Automatiser grupperinger,** beregninger og kalkulasjoner: bruk verktøy og biblioteker med åpen kildekode for automatisering
- **Økosystem og COE:** Bygg opp et internt team for å danne et enhetlig dataøkosystem i selskapet
- **Kultur og læring:** gå bort fra HiPPO -løsninger til datadrevne løsninger
- **Piloter, tilbakemeldinger og oppskalering:** Handle iterativt: Test ut nye metoder i begrenset skala, innhent gyldige tilbakemeldinger, og skaler opp løsningene gradvis.
- **Åpne formater:** Bruk universelle og åpne formater for å være uavhengig av programvareleverandører

- **Maskinlæring:** integrer ML-algoritmer i prosesser for prediksjon og optimalisering
- **IoT og den digitale byggeplassen:** integrer moderne datainnsamlingsteknologi i prosessene
- **Strategisk tilpasning:** forbered deg på fremtidige endringer i bransjen

Det viktigste er å huske at "data alene forandrer ikke et selskap: Det er menneskene som vet hvordan de skal jobbe med dataene, som forandrer det". Legg vekt på kultur, transparente prosesser og en forpliktelse til kontinuerlig forbedring

En systemtilnærming muliggjør et skifte fra digitale siloinitiativer til en fullverdig datadrevet styringsmodell der beslutninger ikke er basert på intuisjon eller antakelser, men på data, fakta og matematisk beregnede sannsynligheter. Den digitale transformasjonen av byggebransjen handler ikke bare om å ta i bruk teknologi, men om å skape et forretningsøkosystem der prosjektinformasjon overføres sømløst og iterativt mellom ulike systemer. Maskinlæringsalgoritmer sørger for automatisk, kontinuerlig analyse, prognosenter og prosessoptimalisering. I et slikt miljø blir spekulasjoner og skjulte data irrelevante - det eneste som gjenstår, er velprøvde modeller, transparente beregninger og forutsigbare resultater.

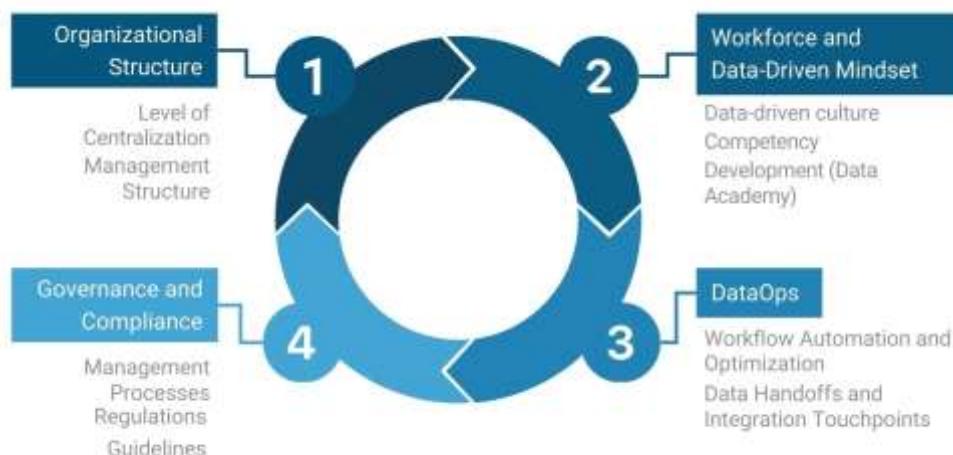


Fig. 10.2-3 Nøkkelelementer for vellykket datahåndtering på bedriftsnivå.

Hver del av boken tilsvarer et bestemt trinn i databehandling og analyse i byggeprosjekter (fig. 2.2-5). Hvis du ønsker å gå tilbake til et av temaene som er diskutert tidligere, og se på det ut fra en helhetlig forståelse av dataflyten, kan du se på deltitlene i fig. 10.2-4.

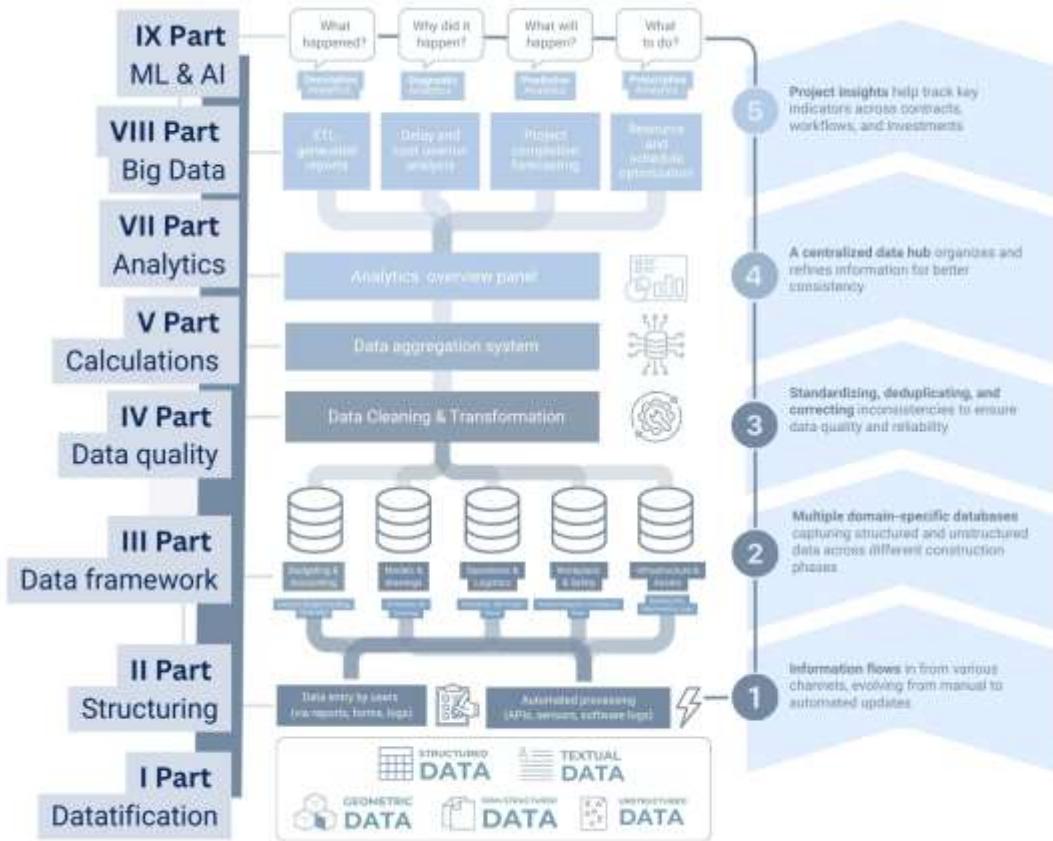


Fig. 10.2-4 Deler av boken i sammenheng med databehandlingspipelinen (fig. 2.2-5): fra digitalisering av informasjon til analyse og kunstig intelligens.

Uansett organisasjonens størrelse, teknologimodenhet eller budsjett, kan du begynne å bevege deg i retning av en datadrevet -tilnærming allerede i dag. Selv små skritt i riktig retning vil gi resultater over tid.

Datadrevet transformasjon er ikke et engangsprosjekt, men en kontinuerlig, iterativ forbedringsprosess som omfatter innføring av nye verktøy, redesign av prosesser og utvikling av en kultur for datadrevet beslutningstaking.

Bygge i Industri 5.0: Hvordan tjene penger når du ikke kan gjemme deg lenger

I lang tid har byggefirmaer tjent penger på ugjennomsiktige prosesser. Den viktigste forretningsmodellen var spekulasjon - overestimering av materialkostnader, arbeidsomfang og prosentvise påslag i lukkede ERP - og PMIS - systemer som er utilgjengelige for eksterne revisorer. Den begrensede tilgangen kundene og deres tillitspersoner hadde til de opprinnelige prosjektdataene, skapte grobunn for ordninger der det ble nesten umulig å verifisere påliteligheten av beregningene.

Denne modellen er imidlertid i ferd med å miste sin relevans. Med demokratiseringen av datatilgang, fremveksten av LLM-er, åpne data og ETL-automatiseringsverktøy er bransjen i ferd med å gå over til en ny arbeidsstandard.

Resultatet er at ugjennomsiktighet ikke lenger er et konkurransefortrinn - snart blir det en byrde som det blir vanskelig å kvitte seg med. Åpenhet går fra å være et alternativ til å bli en forutsetning for å holde seg i markedet.

Hvem vil kundene - banker, investorer, fysiske kunder, private equity, offentlige kunder - samarbeide med i den nye digitale virkeligheten? Svaret er åpenbart: med dem som ikke bare kan levere resultatet, men også begrunnelsen for hvert skritt på veien dit. Etter hvert som mengden åpne data vokser, vil partnere og kunder velge selskaper som garanterer åpenhet, nøyaktighet og forutsigbare resultater.

På bakgrunn av dette vokser det frem nye forretningsmodeller som er basert på datahåndtering og tillit snarere enn spekulasjon:

■ **Selge prosesser i stedet for kvadratmeter:** Den viktigste ressursen blir tillit og effektivitet i stedet for rabatterte konkrete avtaler. Hovedverdien vil være forutsigbarhet i resultatet basert på pålitelige og verifiserte data. Moderne selskaper vil ikke selge byggeobjektet som sådan, men:

- nøyaktige tidsfrister og transparente arbeidsplaner;
- rimelige estimater, underbygget av beregninger;
- full digital sporbarhet og kontroll i alle faser av prosjektet.

■ **Engineering og analyse som en tjeneste:** "Data-as-a-Service"-modellen (en måte å levere ferdige data til brukerne via Internett, som en tjeneste), der hvert prosjekt blir en del av en digital datakjede, og forretningsverdien ligger i evnen til å administrere denne kjeden. Selskapene forvandles til intelligente plattformer som tilbyr løsninger basert på automatisering og analyse:

- Automatisert og transparent utarbeidelse av estimater og planer;
- risiko- og tidsvurdering basert på maskinlæringsalgoritmer;
- beregning av miljøindikatorer (ESG, CO₂, energieffektivitet);
- generering av rapporter fra reviderte åpne kilder.

■ **Produktivisering av ingeniørkompetanse:** Selskapets utvikling kan brukes gjentatte ganger i selskapet og distribueres som et eget produkt - noe som utgjør en ekstra inntektskilde gjennom digitale tjenester. I det nye miljøet skaper selskaper ikke bare prosjekter, men også digitale eiendeler:

- biblioteker med komponenter og estimatmaler;

- automatiserte verifiseringsmoduler;
 - Plugins og skript med åpen kildekode (salg av konsulenttjenester) for arbeid med data.
- **En ny type selskap: den datadrevne integratoren:** en markedsaktør som ikke er avhengig av spesifikke programvareleverandører eller modulære systemer, og som ikke er "låst" til et enkelt programvaregrensesnitt. Det opererer fritt med data - og bygger sin konkurransekraft på dem. Fremtidens byggefirma er ikke bare en entreprenør, men en informasjonsintegrator som kan utføre følgende funksjoner for kunden:
- Kombiner data fra ulike kilder og utfør analyser;
 - Sikre åpenhet og troverdighet i prosessene;
 - gi råd om optimalisering av forretningsprosesser;
 - utvikle verktøy som fungerer i økosystemet for åpne data, LLM, ETL og Pipelines.

Industri 5.0 (fig. 2.1-12) markerer slutten på "de manuelle gjennomsnittsberegningenes tid" og kveldsmøtene mellom administrerende direktører og kalkulasjons- og regnskapsavdelingen. Alt som tidligere var skjult - kalkyler, estimater, volumer - blir åpent, etterprøvbart og forståelig også for ikke-ekspertar. De som er først ute med å omstille seg, blir vinnerne. Alle andre vil bli stående utenfor den nye digitale økonomien i byggebransjen.

KONKLUSJON

Byggebransjen er på vei inn i en tid med grunnleggende endringer. Fra de første opptegnelsene på leirtavler til de enorme mengdene digitale data som strømmer fra prosjektservere og byggeplasser, har informasjonshåndteringen i byggebransjen alltid gjenspeilet den teknologiske modenheten i sin samtid. I dag, med automatisering, åpne formater og intelligente analysesystemer, står bransjen ikke overfor en gradvis utvikling, men en rask digital transformasjon.

Som i andre deler av økonomien må også byggebransjen tenke nytt, ikke bare når det gjelder verktøyene, men også prinsippene for arbeidet. Selskaper som tidligere dikterte markedet og fungerte som det viktigste mellomleddet mellom kunden og prosjektet, er i ferd med å miste sin unike posisjon. Tillit og evnen til å arbeide med data - fra innsamling og strukturering til analyse, prognosering og automatisering av beslutninger - kommer i forgrunnen.

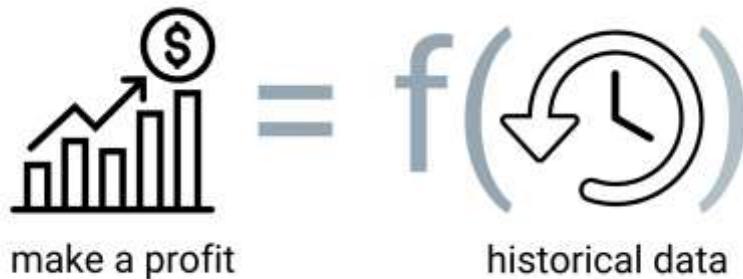


Fig. 10.2-1 Strukturerte historiske data er drivstoffet for en effektiv og oversiktlig virksomhet.

Denne boken beskriver de viktigste prinsippene for datahåndtering i byggebransjen - fra revisjon og standardisering til prosessautomatisering, bruk av visualiseringsverktøy og implementering av intelligente algoritmer. Vi har sett på hvordan du selv med begrensete ressurser kan bygge opp en fungerende dataarkitektur og begynne å ta beslutninger basert på verifiserte fakta i stedet for intuisjon. Arbeid med data er ikke lenger bare en oppgave for IT-avdelingen - det blir grunnlaget for ledelseskulturen, som er avgjørende for selskapets fleksibilitet, tilpasningsevne og langsigtsbærekraft.

Anvendelsen av maskinlæringsteknologi, automatiske prosesseringssystemer, digitale tvillinger og åpne formater gjør det allerede i dag mulig å eliminere den menneskelige faktoren der den tidligere var avgjørende. Bygg- og anleggsbransjen beveger seg mot autonomi og kontrollerbarhet, der bevegelsen fra idé til prosjektrealisering kan sammenlignes med navigering i autopilotmodus: uten avhengighet av subjektive beslutninger, uten behov for manuell inngrisen i alle faser, men med full digital sporbarhet og kontroll (fig. 10.2-2).

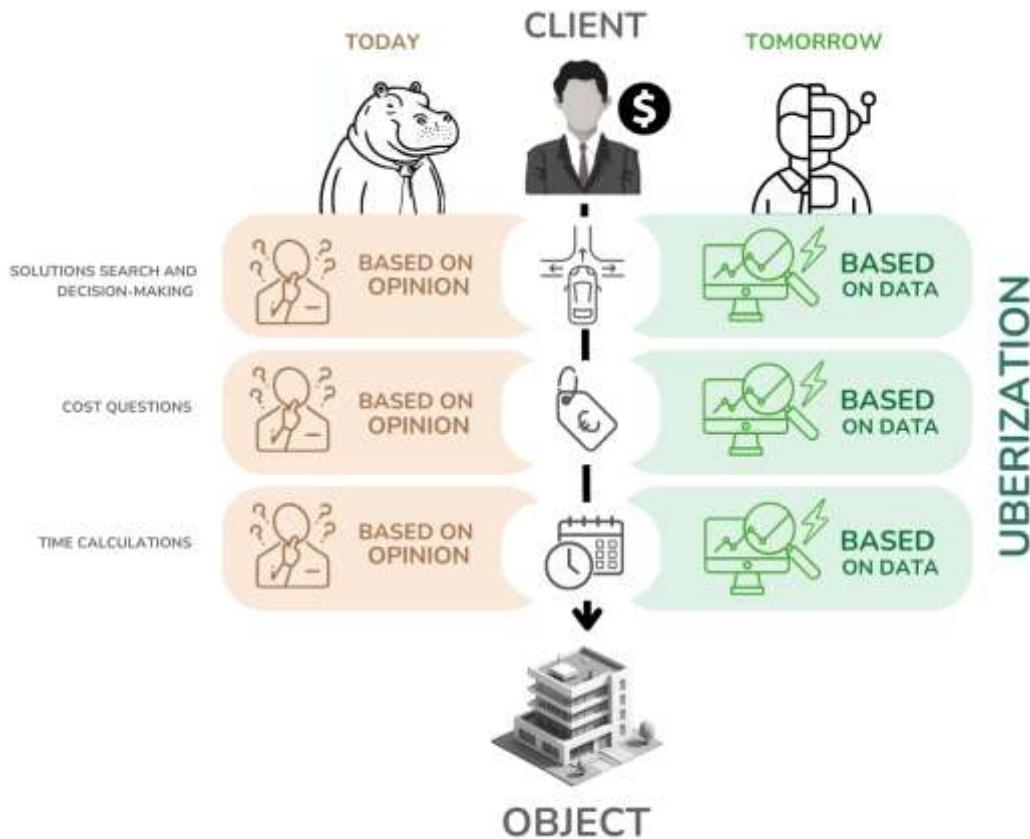


Fig. 10.2-2 Overgangen fra beslutningstaking basert på viktige eksperters meninger (HiPPO) til dataanalyse vil først og fremst bli fremmet av kunden.

Ved å lære deg metodene, prinsippene og verktøyene som presenteres i denne boken, vil du kunne begynne å ta datadrevne, snarere enn intuitive, beslutninger i bedriften din. Du vil også kunne kjøre modulkjeder i LLM-er, kopiere ferdige ETL Pipelines til utviklingsmiljøet ditt (IDE) og automatisk behandle data for å få den informasjonen du trenger, i den formen du ønsker. Med utgangspunkt i bokens kapitler om stordata og maskinlæring vil du deretter kunne implementere mer komplekse scenarier - trekke ut ny innsikt fra historiske data og bruke maskinlæringsalgoritmer til å forutsi og optimalisere prosessene dine.

Åpne data og prosesser vil gi grunnlag for mer nøyaktige estimerater av prosjektkostnader og tidsfrister, slik at byggefirmaene slipper å spekulere på ugjennomsiktige data. Dette er både en utfordring og en mulighet for bransjen til å revurdere sin rolle og tilpasse seg et nytt miljø der åpenhet og effektivitet vil bli viktige suksessfaktorer.

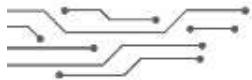
Viljen til å ta kunnskapen til seg og omsette den i praksis er nøkkelen til suksess i den digitale transformasjonens tidsalder.

De bedriftene som først innser dette, vil ha et fortrinn i den nye digitale konkurransen. Men det er viktig å være klar over at data alene ikke vil endre noe som helst. Mange mennesker må endre måten de tenker på, og det krever et incentiv. Bedriften din må tenke nytt om hvordan den deler data.

De som forandrer selskapet, er de som vet hvordan de skal arbeide med disse dataene, tolke dem, bruke

dem til å optimalisere dem og skape en ny prosessarkitektur basert på dem.

Hvis du leser disse linjene, er du klar for forandring, og du er allerede et skritt foran. Takk for at du har valgt denne veien. Velkommen til den digitale transformasjonens tidsalder!



OM FORFATTEREN

Mitt navn er Artyom Boiko. Min reise på byggeplassen startet i 2007 - med en jobb som gruvearbeider i en oljeskifergruve i hjembyen min, mens jeg studerte ved gruveuniversitetet i St. På baksiden av denne boken kan du se en eksplosjonist i gruveveggen der vi utvant og sprengte hundrevis av kubikkmeter med oljeskifer. Karrieren min har tatt meg i mange forskjellige retninger, fra å jobbe som gruve- og underjordisk anleggsarbeider til industriklatrer, taktekker og installatør av heisutstyr. Jeg har hatt æren av å være involvert i prosjekter av ulik skala, fra bygging av private hus til store industrianlegg i ulike deler av verden.



Over tid har arbeidet mitt beveget seg fra fysisk konstruksjon til informasjonshåndtering og digitale prosesser. Siden 2013 har jeg jobbet i ulike stillinger i små, mellomstore og store byggfirmaer i flere regioner i Tyskland, fra designer til datahåndteringssjef. Når det gjelder datahåndtering, har jeg erfaring med å jobbe med data i ulike ERP-systemer, CAD (BIM), MEP, FEM, CMS. Jeg har vært involvert i optimalisering, prosessautomatisering, samt analyse, maskinlæring og databehandling i planleggings-, beregnings- og utførelsesfasene av byggearbeider i industri-, bolig-, infrastruktur- og forsyningsbyggfirmaer.

Jeg har jobbet med programvare med åpen kildekode og åpne data siden 2003. I løpet av denne tiden har jeg implementert mange webprosjekter - fra nettsteder og nettbutikker til fullverdige webapplikasjoner - ved hjelp av åpen kildekode og åpne CMS-løsninger. Disse plattformene, som på mange måter ligner på moderne ERP-systemer, har modulær arkitektur, høy tilpasningsevne og tilgjengelighet. Denne erfaringen la grunnlaget for min profesjonelle tilnærming - et fokus på åpen kildekode-teknologi og en kultur for samarbeidsbasert utvikling. Respekt for åpen kildekode og fri utveksling av kunnskap er noe jeg forsøker å fremme i byggebransjen. Arbeidet mitt for å forbedre datatilgjengeligheten i byggebransjen har resultert i opprettelsen av flere nettsamfunn på sosiale medier der man diskuterer datatilgjengelighet og bruk av åpen kildekode i byggebransjen, samt lanseringen av flere oppstartsbedrifter som utvikler løsninger for å gi tilgang til data fra ulike lukkede systemer og plattformer.

Mitt bidrag til fagmiljøet består i å delta som foredragsholder på konferanser som dekker CAD-interoperabilitet (BIM), ERP, 4D-5D, LLM Machine Learning og Artificial Intelligence, samt artikler publisert i europeiske publikasjoner for byggebransjen. En av mine bemerkelsesverdige prestasjoner er etableringen av "BIM History" [111], et omfattende kart over viktige programvareløsninger for datahåndtering i byggebransjen. Min artikkelserie i sju deler, "BIM Development and Lobbying Games", som er oversatt til flere språk, har fått bred anerkjennelse som et forsøk på å belyse den skjulte dynamikken i digitale standarder.

Det var slik jeg gikk fra å utvinne stein - til å utvinne og systematisere byggdata. Jeg er alltid åpen for faglig dialog, nye ideer og felles prosjekter. Jeg tar gjerne imot tilbakemeldinger og ser frem til å motta meldinger fra deg eller se deg blant mine følgere i sosiale medier. Tusen takk for at du leser denne boken til slutt! Det ville glede meg om denne boken hjelper deg til å forstå temaet data i byggebransjen bedre.

VIDERE FORHOLD

Lesernes meninger spiller en viktig rolle i den videre utviklingen av publikasjonene og i utvelgelsen av prioriterte temaer. Kommentarer om hvilke ideer som har vist seg nyttige, og hvilke som har vakt tvil og krever ytterligere avklaring eller kildehenvisninger, er spesielt verdifulle. Boken inneholder et bredt spekter av materiale og analyser, hvorav noen kan virke kontroversielle eller subjektive. Hvis du i løpet av lesningen finner unøyaktigheter, feilaktige kildehenvisninger, logiske inkonsekvenser eller skrivefeil, er jeg takknemlig for kommentarer, tanker eller kritikk, som du kan sende til:boikoartem@gmail.com. Eller via meldinger på LinkedIn:
linkedin.com/in/boikoartem

Jeg setter pris på all omtale av boken Data-Driven Construction i sosiale medier - deling av leseopplevelser bidrar til å spre budskapet om åpne data og verktøy og støtter arbeidet mitt.

OVERSETTELSESKOMMENTAR

Denne boken er oversatt ved hjelp av kunstig intelligens. Dette har gjort oversettelsesprosessen mye raskere. Men som med all annen teknologi kan det forekomme feil eller unøyaktigheter. Hvis du oppdager noe som ser ut til å være feil eller feilaktig oversatt, vennligst send meg en e-post. Dine kommentarer vil bidra til å forbedre kvaliteten på oversettelsen.

DATADREVNE BYGGESAMFUNN

Det er et sted der du fritt kan stille spørsmål og dele problemer og løsninger:

DataDrivenConstruction.io: <https://datadrivenconstruction.io>

LinkedIn: <https://www.linkedin.com/company/datadrivenconstruction/>

Twitter: <https://twitter.com/datadrivenconst>

Telegram: <https://t.me/datadrivenconstruction>

YouTube: <https://www.youtube.com/@datadrivenconstruction>

ANDRE FERDIGHETER OG KONSEPTER

I tillegg til de viktigste prinsippene for arbeid med data i byggebransjen, tar DataDrivenConstruction for seg en lang rekke andre konsepter, programmer og ferdigheter som er essensielle for den datadrevne profesjonelle. Noen av disse presenteres kun i en oversikt, men er av avgjørende betydning for praksis.

Den interesserte leser kan besøke nettstedet DataDrivenConstruction.io for å finne lenker til ytterligere materiale om nøkkelferdigheter. Dette materialet omfatter arbeid med Python og Pandas, bygging av ETL - prosesser, eksempler på databehandling i CAD-prosjekter i bygg- og anleggsbransjen, stordatasystemer og moderne tilnærminger til visualisering og analyse av bygg- og anleggsdata.

Mange verktøy og programvare med åpen kildekode ble brukt i utarbeidelsen av boken "DataDrivenConstruction" og alle casestudiene. Forfatteren vil gjerne takke utviklerne og medforfatterne av følgende løsnin-

ger:

- Python og Pandas - ryggraden i datamanipulering og automatisering
- Scipy, NumPy, Matplotlib og Scikit-Learn - biblioteker for dataanalyse og maskinlæring
- SQL og Apache Parquet - verktøy for lagring og behandling av store mengder byggdata
- Open Source CAD (BIM) åpne dataverktøy i åpne formater
- N8n, Apache Airflow, Apache NiFi - systemer for orkestrering og automatisering av arbeidsflyt
- DeepSeek, LlaMa, Mistral - Open Source LLM

En spesiell takk til alle deltakerne i diskusjonene om åpne data og verktøy i fagmiljøer og sosiale nettverk, som med sin kritikk, sine kommentarer og ideer har bidratt til å forbedre bokens innhold og struktur.

Følg utviklingen av prosjektet på nettstedet DataDrivenConstruction.io, der det ikke bare publiseres oppdateringer og rettelser til boken, men også nye kapitler, veiledninger og praktiske eksempler på hvordan de beskrevne teknikkene kan brukes.

MAKSIMAL BEKVEMMELIGHET MED DEN TRYKTE VERSJONEN

Du holder en gratis digital versjon av **Datadrevet konstruksjon**. For mer praktisk arbeid og rask tilgang til materialet anbefaler vi deg å ta hensyn til den **trykte utgaven**:



Alltid tilgjengelig: Boken i trykt format blir et pålitelig arbeidsverktøy, slik at du raskt kan finne og bruke de riktige visualiseringene og diagrammene i enhver arbeidssituasjon

Illustrasjoner av høy kvalitet: Alle bilder og all grafikk i den trykte utgaven presenteres i maksimal kvalitet

Rask tilgang til informasjon: enkel navigering, mulighet til å lage notater, bokmerker og arbeide med boken hvor som helst.

Ved å kjøpe den trykte versjonen av boken får du et praktisk verktøy for komfortabelt og effektivt arbeid med informasjon: muligheten til raskt å bruke visuelt materiale i daglige oppgaver, raskt finne de nødvendige ordningene og lage notater. I tillegg støtter kjøpet ditt formidling av åpen kunnskap.

Bestill en trykt versjon av boken på: datadrivenconstruction.io/books



UNIK MULIGHET FOR STRATEGISK POSISJONERING

Vi tilbyr deg å plassere reklamemateriell i gratisversjonen av DataDrivenConstruction. Den betalte versjonen av publikasjonen har tiltrukket seg oppmerksomhet fra spesialister fra mer enn 50 land i verden - fra Latin-Amerika til Asia-Stillehavsregionen - det første året etter utgivelsen. For å diskutere individuelle samarbeidsvilkår og få detaljert informasjon om plasseringsmuligheter, vennligst fyll ut tilbakemeldingsskjemaet på den offisielle portalen datadrivenconstruction.io eller skriv til kontaktene som er oppført på slutten av boken.



BOKKAPITLENE ER TILGJENGELIGE PÅ DATADRIVENCONSTRUCTION.IO

Du kan lese kapitler i Datadrevet bygging på nettstedet for Datadrevet bygging, der deler av boken publiseres fortløpende slik at du raskt kan finne den informasjonen du trenger og bruke den i arbeidet ditt. Her finner du også mange andre publikasjoner om lignende emner, samt eksempler på bruksområder og løsninger som kan hjelpe deg med å utvikle ferdighetene dine og bruke data i bygg- og anleggsbransjen.



SISTE VERSJON AV BOKEN LAST NED FRA DET OFFISIELLE NETTSTEDET

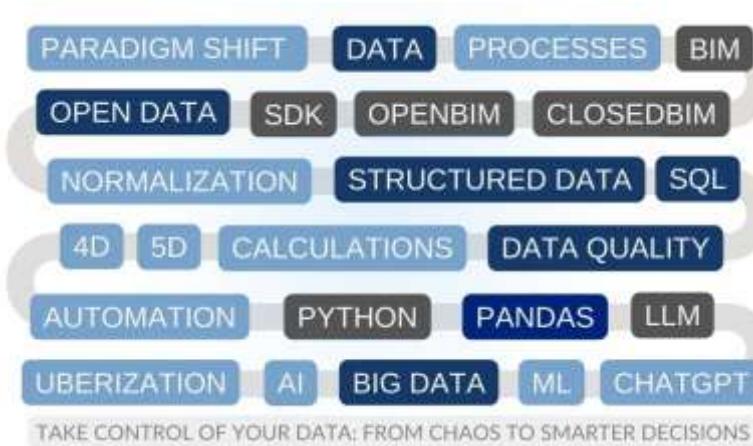
Gjeldende og siste versjon av DataDrivenConstruction-boken er tilgjengelig for nedlasting på datadrivenconstruction.io. Hvis du ønsker å motta oppdateringer med nye kapitler i boken, praktiske tips eller anmeldelser av nye applikasjoner, kan du abonnere på nyhetsbrevet:

- Du vil være den første som blir introdusert til de nye delene av boken
- Få praktiske casestudier og råd om analyse og automatisering i byggebransjen
- Hold deg oppdatert på trender, publikasjoner og eksempler på bruksområder

Gå til datadrivenconstruction.io for å abonnere!

DATADREVET KONSTRUKSJON: RÅDGIVNING, WORKSHOPS OG OPPLÆRING

DataDrivenConstructions opplæringsprogrammer og konsulenttjenester har hjulpet flere titalls ledende byggfirmaer over hele verden med å øke effektiviteten, redusere kostnadene og forbedre kvaliteten på løsningene. DataDrivenConstructions kunder omfatter noen av de største aktørene i milliardmarkedet, inkludert bygg-, konsulent- og IT-selskaper.



Hvorfor velge oss?

- **Relevans:** snakke om de viktigste trendene og innsiktene i bransjen
- **Praktisk:** hjelper fagfolk med å løse hverdagslige oppgaver på en effektiv måte ved hjelp av PoC.
- **Personlig tilnærming:** skreddersydd til din virksomhet, slik at du får maksimalt utbytte av opplæring og rådgivning

DataDrivenConstruction-teamets

hovedfokusområder:

- **Datakvalitetsstyring:** bidra til å parametrisere oppgaver, samle inn krav, validere og klargjøre data for automatisert behandling.
- **Data Mining - utvinning og strukturering av data:** Sett opp ETL-prosesser og hent ut data fra e-post, PDF, Excel, bilder og andre kilder.
- **BIM og CAD analytics:** samle inn, strukturere og analysere informasjon fra RVT-filer, IFC, DWG og andre CAD (BIM)-formater.
- **Analyse og datatransformasjon:** omgjøring av uensartet informasjon til strukturerte data, analyser, innsikt og beslutninger.
- **Dataintegrasjon og prosessautomatisering:** fra automatisk oppretting av dokumenter til integrasjon med interne systemer og eksterne databaser.

Kontakt DataDrivenConstruction.io for å finne ut hvordan automatisering kan hjelpe bedriften din med å oppnå konkrete forretningsresultater.

GLOSSAR

AI (Artificial Intelligence) - **Kunstig intelligens**; datasystemers evne til å utføre oppgaver som normalt krever menneskelig intelligens, for eksempel mønstergjenkjenning, læring og beslutningstaking.

Apache Airflow er en åpen kildekodeplattform for orkestrering av arbeidsflyter som lar deg programmatisk opprette, planlegge og spore arbeidsflyter og ETL-er ved hjelp av DAG-er (directed acyclic graphs).

Apache NiFi er et verktøy for automatisering av dataflyt mellom systemer, med spesialisering innen dataruting og -transformasjon.

Apache Parquet er et effektivt filformat for kolonnelagring av data, optimalisert for bruk i systemer for stordataanalyse. Det gir betydelig komprimering og rask behandling.

API (Application Programming Interface) - et formalisert grensesnitt som gjør det mulig for et program å samhandle med et annet uten tilgang til kildekoden, og som utveksler data og funksjonalitet gjennom standardiserte forespørsler og svar.

Attributt - en egenskap eller egenskap ved et objekt som beskriver dets egenskaper (f.eks. areal, volum, pris, materiale).

Databaser er organiserte strukturer for lagring, administrasjon og tilgang til informasjon, som brukes til effektiv gjenfinning og behandling av data.

BEP (BIM Execution Plan) - En implementeringsplan for bygningsinformasjonsmodellering som definerer målene, metodene og prosessene for implementering av BIM i et prosjekt.

Big Data - store mengder informasjon med betydelig volum, variasjon og oppdateringshastighet, som krever spesiell teknologi for behandling og analyse.

BI (Business Intelligence) - **Business Intelligence**; prosessene, teknologiene og verktøyene som brukes til å omdanne data til meningsfull informasjon for beslutningstaking.

BIM (Building Information Modeling) - Bygningsinformasjonsmodellering; prosessen med å skape og administrere digitale representasjoner av de fysiske og funksjonelle egenskapene til byggeprosjekter, inkludert ikke bare 3D-modeller, men også informasjon om ytelse, materialer, tidsplaner og kostnader.

BlackBox/WhiteBox - tilnæringer til å forstå systemet: I det første tilfellet er den interne logikken skjult, bare innganger og utganger er synlige; i det andre tilfellet er behandlingen transparent og tilgjengelig for analyse.

Bounding Box er en geometrisk konstruksjon som beskriver grensene for et objekt i et tredimensjonalt rom ved hjelp av minimums- og maksimumskoordinatene X, Y og Z, slik at det dannes en "boks" rundt objektet.

BREP (Boundary Representation) er en geometrisk representasjon av objekter som definerer dem gjennom grensene til overflater.

CAD (Computer-Aided Design) er et datastøttet designsystem som brukes til å lage, redigere og analysere nøyaktige tegninger og 3D-modeller innen arkitektur, bygg og anlegg, ingeniørfag og andre bransjer.

CAFM (Computer-Aided Facility Management) er en programvare for eiendoms- og infrastrukturforvaltning som omfatter arealplanlegging, kapitalforvaltning, vedlikehold og kostnadsovervåking.

CDE (Common Data Environment) - et sentralisert digitalt område for håndtering, lagring, deling og samarbeid med prosjektinformasjon i alle faser av anleggets livssyklus.

Et Centre of Excellence (CoE) er en spesialisert struktur i en organisasjon som er ansvarlig for å utvikle et spesifikt kunnskapsområde, utvikle standarder og beste praksis, lære opp ansatte og støtte innføringen av innovasjoner.

CoClass er et moderne tredjegenerasjons klassifiseringssystem for bygningselementer.

En konseptuell datamodell er en høynivårepresentasjon av grunnleggende entiteter og deres relasjoner uten attributdetaljer, som brukes i de innledende fasene av databasedesignet.

CRM (Customer Relationship Management) er et system for styring av kunderelasjoner som brukes til å automatisere salgs- og serviceprosesser.

DAG (Directed Acyclic Graph) er en rettet acyklistisk graf som brukes i dataorkestreringssystemer (Airflow, NiFi) for å bestemme oppgaverekkefølger og avhengigheter.

Dash er et Python-rammeverk for å lage interaktive nettbaserte datavisualiseringer.

Dashbord - Et dashbord som visuelt presenterer viktige ytelsesindikatorer og måltall i sanntid.

Den datasentriske tilnærmingen er en metodikk som prioriterer data fremfor applikasjoner eller programvarekode, og som gjør data til organisasjonens sentrale ressurs.

Datastyring - et sett med rutiner, prosesser og retningslinjer som sikrer hensiktsmessig og effektiv bruk av data i en organisasjon, inkludert tilgangs-, kvalitets- og sikkerhetskontroller.

Data Lake er et lagringsanlegg som er utviklet for å lagre store mengder rådata i sitt opprinnelige format til de skal brukes.

Data Lakehouse er en arkitektonisk tilnærming som kombinerer fleksibiliteten og skalerbarheten til datasjøer (Data Lake) med håndterbarheten og ytelsen til datavarehus (DWH).

Datadrevet bygging er en strategisk tilnærming der hvert trinn i anleggets livssyklus - fra prosjektering til drift - støttes av automatiserte, sammenkoblede systemer. Denne tilnærmingen gir kontinuerlig, faktabasert læring, reduserer usikkerheten og gjør det mulig for selskaper å oppnå et bærekraftig lederskap i bransjen.

Data-Driven Integrator - et selskap som spesialiserer seg på å kombinere data fra ulike kilder og analysere dem for å ta ledelsesbeslutninger.

Datadrevet tilnærming - en metodikk der data blir sett på som en strategisk ressurs, og beslutninger tas basert på objektiv analyse av informasjon i stedet for subjektive meninger.

Dataminimalisme - en tilnærming som går ut på å redusere data til det mest verdifulle og meningsfulle, noe som muliggjør forenklet behandling og analyse av informasjon.

Datasump - En masse ustrukturerte data som oppstår når informasjon samles inn og lagres på en ukontrollert måte uten skikkelig organisering.

DataOps er en metodikk som kombinerer DevOps-prinsipper, data og analyse, med fokus på å forbedre samarbeid, integrasjon og automatisering av datastrømmer.

Digitalisering av informasjon er prosessen med å konvertere alle aspekter ved byggevirksomhet til en digital form som er egnet for analyse, tolkning og automatisering.

DataFrame - En todimensjonal tabellformet datastruktur i Pandas-biblioteket, der rader representerer individuelle poster eller objekter, og kolonner representerer deres egenskaper eller attributter.

Deskriptiv analyse - analyse av historiske data for å forstå hva som har skjedd i fortiden.

Diagnostisk analyse - Analyse av data for å finne ut hvorfor noe skjedde.

Et Gantt-diagram er et prosjektplanleggingsverktøy som representerer oppgaver som horisontale søyler på en tidslinje, slik at du kan visualisere rekkefølgen og varigheten av arbeidet.

DWH (Data Warehouse) er et sentralisert datavarehussystem som samler informasjon fra flere kilder, strukturerer den og gjør den tilgjengelig for analyse og rapportering.

ESG (Environmental, Social, Governance) - et sett med kriterier for å vurdere de miljømessige, sosiale og styringsmessige konsekvensene av et selskap eller et prosjekt.

ELT (Extract, Load, Transform) er en prosess der data først hentes ut fra kilder og lastes inn i et datalager, og deretter transformeres for analytiske formål.

ETL (Extract, Transform, Load) er prosessen med å hente ut data fra ulike kilder, transformere dem til ønsket format og laste dem inn i mållageret for analyse.

ER-diagram (Entity-Relationship) - et visuelt diagram som viser entiteter, deres attributter og relasjonene mellom dem, og som brukes i datamodellering.

ERP (Enterprise Resource Planning) er et omfattende, modulbasert ressursplanleggingssystem som brukes til å administrere og optimalisere ulike aspekter ved byggeprosessen.

Funksjoner - I maskinlæring, uavhengige variabler eller attributter som brukes som inndata til en modell.

Fysisk datamodell - en detaljert fremstilling av databasestrukturen, inkludert tabeller, kolonner, datatyper, nøkler og indekser, optimalisert for et bestemt DBMS.

FPDF er et Python-bibliotek for oppretting av PDF-dokumenter.

Geometric Core er en programvarekomponent som inneholder grunnleggende algoritmer for oppretting, redigering og analyse av geometriske objekter i CAD, BIM og andre tekniske applikasjoner.

HiPPO (Highest Paid Person's Opinion) - en tilnærming til beslutningstaking som baserer seg på oppfatningen til den høyest betalte personen i organisasjonen i stedet for objektive data.

IDE (Integrated Development Environment) - integrert utviklingsmiljø, et omfattende verktøy for skriving, testing og feilsøking av kode (f.eks. PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) er en spesifikasjon for informasjonsleveranser som definerer datakravene i ulike faser av et prosjekt.

IFC (Industry Foundation Classes) er et format for utveksling av BIM-data som sikrer interoperabilitet mellom ulike programvareløsninger.

Industri 5.0 er et industrielt utviklingskonsept som kombinerer mulighetene som ligger i digitalisering, automatisering og kunstig intelligens med menneskelig potensial og miljømessig bærekraft.

Dataintegrasjon er prosessen med å kombinere data fra ulike kilder til ett enkelt, sammenhengende system for å gi en enhetlig oversikt over informasjonen.

Informasjonssiloer er isolerte datalagringssystemer som ikke deler informasjon med andre systemer, noe som skaper hindringer for effektiv bruk av data.

IoT (Internet of Things) er konseptet som går ut på å koble fysiske gjenstander til Internett for å samle inn, behandle og overføre data.

k-NN (k-Nearest Neighbors) er en maskinlæringsalgoritme som klassifiserer objekter basert på likheten med de nærmeste naboene i treningseksemplet.

Kaggle er en plattform for dataanalyse og maskinlæringskonkurranser.

Kalkulasjon - beregning av kostnadene for byggearbeider eller -prosesser for en bestemt måleenhet (f.eks. 1 m² gipsplatevegg, 1 m³ betong).

KPI-er (Key Performance Indicators) er nøkkelindikatorer, kvantifiserbare måltall som brukes til å evaluere suksessen til et selskap eller et bestemt prosjekt.

Etiketter - I maskinlæring, målvariablene eller attributtene som modellen skal forutsi.

Læringsalgoritme - Prosessen med å finne den beste hypotesen i en modell som svarer til en målfunksjon ved hjelp av et sett med treningsdata.

Lineær regresjon - En statistisk metode for modellering av forholdet mellom en avhengig variabel og én eller flere uavhengige variabler.

LLM (Large Language Model) - Large Language Model, en kunstig intelligens som er opplært til å forstå og generere tekst fra store datamengder, og som er i stand til å analysere kontekst og skrive programmeringskode.

LOD (Level of Detail/Development) - detaljnivået i modellen som bestemmer graden av geometrisk nøyaktighet og informasjonsinnhold.

En logisk datamodell er en detaljert beskrivelse av entiteter, attributter, nøkler og relasjoner som gjenspeiler forretningsinformasjon og -regler, og er et mellomstadium mellom den konseptuelle og den fysiske modellen.

Maskinlæring - En klasse av kunstig intelligens-teknikker som gjør det mulig for datasystemer å lære og forutsi ut fra data uten eksplisitt programmering.

Masterformat er et førstegenerasjons klassifiseringssystem som brukes til å strukturere byggespesifikasjoner i seksjoner og fagområder.

MEP (Mechanical, Electrical, Plumbing) - Bygningstekniske systemer som omfatter mekaniske, elektriske og VVS-relaterte komponenter.

Mesh er en nettrepresentasjon av 3D-objekter som består av hjørner, kanter og flater.

Modell - I maskinlæring et sett med ulike hypoteser, hvorav én tilnærmer seg målfunksjonen som skal predikeres eller tilnærmes.

Datamodellering er prosessen med å skape en strukturert representasjon av data og deres relasjoner for implementering i informasjonssystemer, inkludert konseptuelle, logiske og fysiske nivåer.

n8n er et åpen kildekode-verktøy for automatisering av arbeidsflyter og integrering av applikasjoner ved hjelp av en lavkodet tilnærming.

Normalisering - innen maskinlæring, prosessen med å bringe ulike numeriske data til en felles skala for å gjøre det enklere å behandle og analysere dem.

Reverse engineering - prosessen der man studerer et objekts enhet, funksjon og produksjonsteknologi ved å analysere dets struktur, funksjoner og virkemåte. I datasammenheng - å hente ut informasjon fra proprietære formater for bruk i åpne systemer.

OCR (Optical Character Recognition) er en teknologi for **optisk tegngjenkjenning** som konverterer tekstsbilder (skannede dokumenter, bilder) til et maskinlesbart tekstformat.

OmniClass er andre generasjons internasjonale klassifiseringsstandard for håndtering av bygginformasjon.

Ontologi - Et system av begreper som formaliserer et bestemt kunnskapsfelt.

Open Source - en modell for utvikling og distribusjon av programvare med åpen kildekode som er tilgjengelig for fri bruk, studier og endringer.

Åpen BIM er et konsept som innebærer bruk av åpne standarder og formater for datautveksling mellom ulike programvareløsninger.

Åpne standarder - offentlig tilgjengelige spesifikasjoner for en bestemt oppgave som gjør det mulig for ulike systemer å samhandle og utveksle data.

Pandas er et Python-bibliotek med åpen kildekode for databehandling og -analyse, og tilbyr DataFrame- og Series-datastrukturer for effektiv håndtering av tabellinformasjon.

Det åpne dataparadigmet er en tilnærming til databehandling der informasjon gjøres fritt tilgjengelig for bruk, gjenbruk og spredning av hvem som helst.

Parametrisk metode er en estimeringsmetode for byggeprosjekter som bruker statistiske modeller for å estimere verdien basert på prosjektparametere.

PIMS (Project Information Model) er et digitalt system som er utviklet for å organisere, lagre og dele all prosjektinformasjon.

Pipeline - En sekvens av databehandlingsprosesser, fra uthenting og transformasjon til analyse og visualisering.

PMIS (Project Information Management System) er et prosjektstyringssystem som er utviklet for detaljert oppgavekontroll på det enkelte byggeprosjektnivå.

Prediktiv analyse er en del av analyse som bruker statistiske metoder og maskinlæring til å forutsi fremtidige resultater basert på historiske data.

Prescriptive Analytics - En del av analysen som ikke bare forutser fremtidige resultater, men som også foresår optimale tiltak for å oppnå de ønskede resultatene.

Proprietære formater er lukkede dataformater som kontrolleres av et bestemt selskap, noe som begrenser muligheten til å dele informasjon og øker avhengigheten av spesifikk programvare.

QTO (Quantity Take-Off) er prosessen med å hente ut mengdene av elementer fra designdokumentene for å beregne materialmengdene som kreves for å realisere prosjektet.

Kvalitetsstyringssystem - et kvalitetsstyringssystem som sikrer at prosesser og resultater oppfyller fastsatte krav.

RAG (Retrieval-Augmented Generation) er en metode som kombinerer de generative egenskapene til språkmodeller med utvinning av relevant informasjon fra bedriftsdatabase, noe som forbedrer nøyaktigheten og relevansen av svarene.

RDBMS (Relational Database Management System) er et relasjonsdatabasestyringssystem som organiserer informasjon i form av sammenhengende tabeller.

RegEx (Regular Expressions) er et formalisert språk for søk og behandling av strenger, som gjør det mulig å spesifisere maler for kontroll av tekstdata for å se om de oppfyller bestemte kriterier.

Regresjon er en statistisk metode for å analysere forholdet mellom variabler.

CO₂-beregninger er en metode for å estimere karbondioksidutslipp knyttet til produksjon og bruk av byggematerialer og -prosesser.

Ressursmetoden er en metode for å lage estimater basert på en detaljert analyse av alle ressursene (materialer, arbeidskraft, utstyr) som kreves for å utføre byggearbeidene.

RFID (Radio Frequency Identification) er en teknologi for automatisk identifisering av objekter ved hjelp av radiosignaler, og brukes til sporing av materialer, maskiner og personell.

ROI (Return on Investment) er en indikator som gjenspeiler forholdet mellom fortjeneste og investerte midler, og som brukes til å vurdere effektiviteten av investeringer.

SaaS (Software as a Service) er en modell for programvare som en tjeneste, der applikasjonene hostes av en leverandør og gjøres tilgjengelig for brukerne over Internett.

SCM (Supply Chain Management) - styring av forsyningskjeden, som omfatter koordinering og optimalisering av alle prosesser fra innkjøp av materialer til levering av ferdige produkter.

Datasiloer er isolerte informasjonslagre i en organisasjon som ikke er integrert med andre systemer, noe som gjør det vanskelig å dele data og ineffektivt.

SQL (Structured Query Language) er et strukturert spørrespråk som brukes til å arbeide med relasjonsdatabaser.

SQLite er et lett, innebygd og plattformuavhengig DBMS som ikke krever en separat server, og som støtter grunnleggende SQL-funksjoner og er mye brukt i mobile applikasjoner og innebygde systemer.

Strukturerte data - informasjon som er organisert i et bestemt format med en klar struktur, for eksempel i relasjonsdatabaser eller tabeller.

Løst strukturerte data - informasjon med delvis organisering og fleksibel struktur, for eksempel JSON eller XML, der ulike elementer kan inneholde ulike sett med attributter.

En entitet er et konkret eller abstrakt objekt i den virkelige verden som kan identifiseres, beskrives og representeres entydig i form av data.

Overvåket læring - En type maskinlæring der en algoritme trenes opp på oppdelte data der det ønskede resultatet er kjent for hvert eksempel.

Taksonomi er et hierarkisk klassifikasjonssystem som brukes til systematisk å kategorisere elementer basert på fellestrekker.

Titanic Dataset er et populært datasett for opplæring og testing av maskinlæringsmodeller.

Opplæring - Prosessen der en maskinlæringsalgoritme analyserer data for å identifisere mønstre og danne en modell.

Overføringslæring er en maskinlæringssteknikk der en modell som er trent opp for én oppgave, brukes som utgangspunkt for en annen oppgave.

Transformasjon - Prosessen med å endre formatet, strukturen eller innholdet i data for senere bruk.

Datakrav - formaliserte kriterier som definerer struktur, format, fullstendighet og kvalitet på informasjonen som kreves for å støtte forretningsprosessene.

Uberisering av byggebransjen er en prosess der tradisjonelle forretningsmodeller **i byggebransjen** endres under påvirkning av digitale plattformer som gir direkte interaksjon mellom kunder og entreprenører uten mellomledd.

Uniclass er et andre- og tredjegenerasjons klassifiseringssystem for bygningselementer som er mye brukt i Storbritannia.

USD (Universal Scene Description) er et dataformat som er utviklet for datagrafiikk, men som har fått anvendelse i tekniske systemer på grunn av sin enkle struktur og uavhengighet av geometriske kjerner.

Datavalidering er en prosess der man kontrollerer informasjon opp mot etablerte kriterier og krav for å sikre at dataene er nøyaktige, fullstendige og konsistente.

Vektordatabase - En spesialisert type database som lagrer data som flerdimensjonale vektorer for effektivt semantisk søk og sammenligning av objekter.

Vektorrepresentasjon (embedding) er en metode for å omdanne data til flerdimensjonale numeriske vektorer som gjør det mulig for maskinalgoritmer å behandle og analysere informasjon på en effektiv måte.

VectorOps er en metodikk som fokuserer på behandling, lagring og analyse av flerdimensjonale vektordata, noe som er spesielt relevant på områder som digitale tvillinger og semantiske søk.

Visualisering - Grafisk fremstilling av data for bedre oppfatning og analyse av informasjon.

Den alfabetiske kategoriseringen av begrepene ble gjort etter deres engelske navn.

LISTE OVER REFERANSER OG NETTBASERT MATERIALE

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12. desember 2016. [Online]. Tilgjengelig: <https://www.gartner.com/en/documents/3530919>. [Dato for tilgang: 1. mars 2025].
- [2] KPMG, "Kjente utfordringer - nye tilnærmingar. 2023 Global Construction Survey", 1. januar 2023. [Online]. Tilgjengelig: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Besøkt: 5. mars 2025].
- [3] F. R. Barnard, "Et bilde sier mer enn tusen ord", 10 Mapi 1927. [Online]. Tilgjengelig: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Referansedato: 15. mars 2025].
- [4] M. Bastian, "Microsoft CEO Satya Nadella sier at selvpåstått AGI er "meningsløs referansehacking", 21. februar 2025. [Online]. Tilgjengelig: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Dato for tilgang: 15. mars 2025].
- [5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1. januar 2016. [Online]. Available: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Dato for tilgang: 2. mars 2025].
- [6] C. D. Gillespie, "Leire: sammenfiltringen av jorden i leirens tidsalder", 2024. [Online]. Tilgjengelig: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papyrus fra det 3. århundre f.Kr. Språket er gresk", 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Overvåking: bruk av de verktøyene som er tilgjengelige", 1980. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Dato for adresse: 15. mars 2025].
- [9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15. februar 2015. [Online]. Tilgjengelig: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Dato for tilgang: 15. mars 2025].
- [10] Skanska USA, "Fall Construction Market Trends", 2. november 2023. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Oxford Essential Quotations (4. utg.)", Oxford University Press, 2016. [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236>. [Dato for tilgang: 1. mars 2025].

- [12] "Sitat: Sondergaard om dataanalyse," [Online]. Available: <https://www.causeweb.org/cause/resources/library/r2493>. [Dato for tilgang: 15. mars 2025].
- [13] "Hvordan den globale interessen for kunstig intelligens øker markedet for datahåndtering", 28. mai 2024. [Online]. Tilgjengelig: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Dato for tilgang: 15. mars 2025].
- [14] И. McCue, "ERP-historie", 2024. [Online]. Tilgjengelig: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "ERP-pris: Hvor mye koster et ERP-system?", [Online]. Tilgjengelig: <https://erpscout.de/en/erp-costs/>. [Dato for tilgang: 15. mars 2025].
- [16] softwarepath, "Hva 1384 ERP-prosjekter forteller oss om valg av ERP (ERP-rapport 2022)", 18. januar 2022. [Online]. Tilgjengelig: <https://softwarepath.com/guides/erp-report>. [Dato for tilgang: 15. mars 2025].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. desember 2016. [Online]. Tilgjengelig: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Dato for adresse: 1. mars 2025].
- [18] Mckinsey, "The data-driven enterprise of 2025", 28. januar 2022. [Online]. Available: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Referansedato: 22. mai 2024].
- [19] Wikipedia, "Moores lov". [Online]. Tilgjengelig: https://en.wikipedia.org/wiki/Moore%27s_law. [Dato for tilgang: 15. mars 2025].
- [20] Accenture, "Building More Value With Capital Projects", 1. januar 2020. [Online]. Available: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Dato for tilgang: 3. mars 2024].
- [21] B. Marr, "Hvor mye data skaper vi hver dag? Den forbløffende statistikken alle bør lese", 2018. [Online]. Tilgjengelig: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read>.
- [22] "Hvor mye data produseres hver dag?", 2024. [Online]. Tilgjengelig: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] T. Sullivan, "AI og den globale "datasfæren": hvor mye informasjon vil menneskeheten ha innen 2025?", 2024. [Online]. Tilgjengelig: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Totalt antall trykte bøker produsert i ulike regioner i Vest-Europa i hvert halve århundre mellom 1454 og 1800", [Internett]. [Online]. Available:

- <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Referansdato: 1. mars 2025].
- [25] "Eksempler på prising", 2024. [Online]. Tilgjengelig: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Enterprises outsource data storage as complexity rises", 10. mai 2024. [Online]. Tilgjengelig: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Dato for tilgang: 15. mars 2025].
- [27] JETSOFTPRO, "SaaS er død? Microsofts administrerende direktørs sjokkerende spådom forklart", 13. januar 2025. [Online]. Tilgjengelig: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. desember 2024. [Online]. Tilgjengelig: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Dato for tilgang: 15. mars 2025].
- [29] GoodReads, "Tim Berners-Lee", [Online]. Available: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Dato for tilgang: 15. mars 2025].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1. januar 2023. [Online]. Available: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Besøkt: 5. mars 2025].
- [31] I. Deininger, B. Koch, R. Bauknecht og M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and an energy model", 2024. [Online]. Tilgjengelig: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Stud...
- [32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1. februar 2017. [Online]. Tilgjengelig: <https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf>.
- [33] Construction Task Force to the Deputy Prime Minister, "Rethinking Construction", 1. oktober 2014. [Online]. Tilgjengelig: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] Forbes, "Without An Opinion, You're Just Another Person With Data", 15. mars 2016. [Online]. Available: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Dato for tilgang: 15. mars 2025].
- [35] Wikiquote, "Charles Babbage", [Online]. Tilgjengelig: https://en.wikiquote.org/wiki/Charles_Babbage. [Dato for tilgang: 15. mars 2025].

- [36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves", 12. mars 2025. [Online]. Available: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Dato for tilgang: 15. mars 2025].
- [37] The Canadian Construction Association og KPMG i Canada, 2021, "Construction in a digital world", 1. mai 2021. [Online]. Available: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Dato for tilgang: 5. mars 2025].
- [38] ZCS, "Decoding the Fifth Industrial Revolution". [Online]. Tilgjengelig: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Dato for tilgang: 15. mars 2025].
- [39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.
- [40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software", 24. januar 2024. [Online]. Tilgjengelig: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Dato for tilgang: 15. mars 2025].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1. september 2008. [Online]. Tilgjengelig: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Improving construction productivity", [Online]. Tilgjengelig: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Dato for adresse: 15. mars 2025].
- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19. januar 2023. [Online]. Tilgjengelig: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Dato for tilgang: 1. mars 2025].
- [44] McKinsey, "Delivering on construction productivity is no longer optional", 9. august 2024. [Online]. Tilgjengelig: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Dato for tilgang: 5. mars 2025].
- [45] ING Group, "Lagging productivity in construction is driving up construction costs", 12. desember 2022. [Online]. Tilgjengelig: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Dato for adresse: 15. mars 2025].
- [46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software'", 19. desember 2024. [Online]. Tilgjengelig: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Dato for tilgang: 15. mars 2025].
- [47] Business Insider, "Anthropics CEO sier at om 3 til 6 måneder vil AI skrive 90 % av koden programvareutviklere hadde ansvaret for", 15. mars 2025. [Online]. Tilgjengelig:

- <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Dato for tilgang: 30. mars 2025].
- [48] Statista, "Popularitetsammenligning av databasehåndteringssystemer (DBMS) over hele verden per juni 2024, etter kategori," juni 2024. [Online]. Tilgjengelig: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Dato for adresse: 15. mars 2025].
- [49] DB-Engines, "DB-Engines Ranking", [Online]. Available: <https://db-engines.com/en/ranking>. [Referansedato: 15. mars 2025].
- [50] "Stack Overflow Developer Survey 2023", 2024. [Online]. Tilgjengelig: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Tilgjengelig: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Strukturerte og ustukturerte data: Hva er forskjellen?", 2024. [Online]. Tilgjengelig: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF COMPARISON OF DATA FORMATS FOR CONSTRUCTION PROJECTS", 23. april 2024. [Online]. Tilgjengelig: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Building Information Modeling Whitepaper site", 2003. [Online]. Tilgjengelig: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Boyko, "Lobbykrig og BIM-utvikling. Del 5: BlackRock er mesteren over alle teknologier. Hvordan selskaper kontrollerer åpen kildekode", 2024. [Online]. Tilgjengelig: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 14.11.2011. [Online]. Tilgjengelig: https://isicad.net/articles.php?article_num=14805. [Dato for adresse: 02 2025].
- [57] C. Eastman og A. Cthers, "Eastman, Charles; And Cthers", september 1974. [Online]. Tilgjengelig: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Dato for tilgang: 15. mars 2025].
- [58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 11. november 2011. [Online]. Tilgjengelig: https://isicad.net/articles.php?article_num=14805. [Referansedato: 15. mars 2025].
- [59] D. Weisberg, "History of CAD", 12. desember 2022. [Online]. Tilgjengelig: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Referansedato: 15. mars 2025].

- [60] ADSK, "White Paper Building Information Modeling", 2002. [Online]. Tilgjengelig: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Referansedato: 15. mars 2025].
- [61] ADSK, "White Paper Building Information Modeling in Practice", [Online]. Available: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Dato for tilgang: 15. mars 2025].
- [62] A. Boyko, "Lobbykrieg og BIM-utvikling. Del 2: åpen BIM VS lukket BIM. Europa VS resten av verden", 2024. [Online]. Tilgjengelig: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
- [63] A. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs," 2024. [Online]. Tilgjengelig: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "Whitepaper BIM", 2002. [Online]. Tilgjengelig: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Dato for adresse: 15. mars 2025].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale", [Online]. Available: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Dato for tilgang: 15. mars 2025].
- [66] M. Shacklett, "Strukturerte og ustukturerte data: Viktige forskjeller", 2024. [Online]. Tilgjengelig: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Tilgjengelig: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the", 2004. [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [Dato for tilgang: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Tilgjengelig: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Dato for tilgang: 15. mars 2025].
- [70] Analyticsindiamag, "6 mest tidkrevende oppgaver for dataforskere", 15. mai 2019. [Online]. Tilgjengelig: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "Rapport: Dataforskere bruker mesteparten av tiden på å rydde opp", 06. juli 2015. [Online]. Tilgjengelig: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Dato for tilgang: 5. mars 2025].

- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. juli 2015. [Online]. Tilgjengelig: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Dato for tilgang: 15. mars 2025].
- [73] B. Cyphers og K. Doctorow, "Personvern uten monopol: Databeskyttelse og interoperabilitet", 2024. [Online]. Tilgjengelig: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1. oktober 2013. [Online]. Tilgjengelig: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Dato for tilgang: 15. mars 2025].
- [75] A. Boyko, "Kampen for åpne data i byggebransjen. Historien om AUTOLISP, intelliCAD, openDWG, ODA og openCASCADE", 15.05.2024. [Online]. Tilgjengelig: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Dato for adresse: 16 02 2025].
- [76] Wikipedia, "Microsoft og åpen kildekode", [Online]. Tilgjengelig: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Dato for tilgang: 15. mars 2025].
- [77] TIME, "Gapet mellom åpne og lukkede AI-modeller kan krympe. Her er hvorfor det betyr noe", 5. november 2024. [Online]. Tilgjengelig: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Dato for tilgang: 15. mars 2025].
- [78] The Verge, "Mer enn en fjerdedel av ny kode hos Google genereres av AI", 29. oktober 2024. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Dato for tilgang: 15. mars 2025].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15. desember 2020. [Online]. Tilgjengelig: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing>. [Tilgang: 15. mars 2025].
- [80] PwC, "PwC Open Source Monitor 2019", 2019. [Online]. Available: <https://www.pwc.de/de/digital-transformation/open-source-monitor-research-report-2019.pdf>. [Besøkt 15. mars 2025].
- [81] Travers Smith, "Den åpne hemmeligheten: Programvare med åpen kildekode", 2024. [Online]. Tilgjengelig: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Dato for tilgang: 15. mars 2025].
- [82] Deloitte, "The data transfer process in corporate transformations", 2021. [Online]. Tilgjengelig: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Referansedato: 15. mars 2025].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20. mars 2024. [Online]. Available: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery>.

- delivery/data-analytics-and-ai-in-government-project-delivery. [Dato for tilgang: 5. mars 2025].
- [84] "Sitat Origin: Alt skal gjøres så enkelt som mulig, men ikke enklere," 13. mai 2011. [Online]. Available: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Dato for tilgang: 15. mars 2025].
- [85] "Transformer (arkitektur for dyp læring)" [Online]. Tilgjengelig: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Dato for tilgang: 15. mars 2025].
- [86] "Nedlastingsstatistikk for Python-pakker", 2024. [Online]. Tilgjengelig: <https://www.pepy.tech/projects/pandas>.
- [87] Intervju Bit, "Topp 10 Python-biblioteker", 2023. [Online]. Tilgjengelig: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>. [Dato for tilgang: 30. mars 2025].
- [88] "NVIDIA and HP Supercharge Data Science and Generative AI on Workstations", 7. mars 2025. [Online]. Tilgjengelig: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Dato for tilgang: 15. mars 2025].
- [89] P. Orac, "Hvordan behandle en DataFrame med millioner av rader på få sekunder," 2024. [Online]. Tilgjengelig: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Hva er Kaggle?", 2024. [Online]. Tilgjengelig: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "NVIDIAs administrerende direktør Jensen Huang Keynote på COMPUTEX 2024", 2. juni 2024. [Online]. Tilgjengelig: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Besøkt 15. mars 2025].
- [92] "Medlemmer: Stiftere og bedriftsmedlemmer", 2024. [Online]. Tilgjengelig: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts", 24. november 2024. [Online]. Tilgjengelig: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Dato for tilgang: 23. februar 2025].
- [94] A. Boyko, "Verden etter BIM. Overgang til data og prosesser og hvorvidt byggebransjen trenger semantikk, formater og interoperabilitet", 20.12.2024. [Online]. Tilgjengelig: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Dato for adresse: 23. februar 2025].
- [95] N. I. o. Health, "NIH STRATEGISK PLAN FOR DATA SCIENCE", 2016. [Online]. Available:

- https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Dato for tilgang: 23. februar 2025].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 trillion Per Year", 22. september 2016. [Online]. Tilgjengelig: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Impacts of Data Quality", 1. januar 2025. [Online]. Tilgjengelig: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects," [Online]. Available: <https://www.wbdg.org/resources/design-for-maintainability>. [Dato for tilgang: 15. mars 2025].
- [99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", april 2014. [Online]. Available: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%202014.pdf>. [Dato for tilgang: 15. mars 2025].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1. januar 2025. [Online]. Tilgjengelig: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "I mangel av en spiker" [Online]. Tilgjengelig: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Dato for tilgang: 15. mars 2025].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", oktober 2013. [Online]. Tilgjengelig: https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Dato for tilgang: 15. mars 2025].
- [103] EY, "The path to carbon neutrality", 10. mars 2023. [Online]. Available: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Dato for tilgang: 15. mars 2025].
- [104] PWC, "ESG Awareness", 1. juli 2024. [Online]. Tilgjengelig: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Dato for tilgang: 15. mars 2025].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Tilgjengelig: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO₂_beregning av innebygd karbon", 2024. [Online]. Tilgjengelig: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.
- [107] McKinsey, "Imagining Construction's Digital Future", 24. juni 2016. [Online]. Tilgjengelig: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital>

- future. [Dato for tilgang: 25. februar 2025].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10. oktober 2024. [Online]. Tilgjengelig: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Dato for adresse: 15. mars 2025].
- [109] SAS, "Data lake and data warehouse - know the difference," [Online]. Available: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Referansedato: 15. mars 2025].
- [110] ADSK, "Bygningsinformasjonsmodellering", 2002. [Online]. Tilgjengelig: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Dato for adresse: 15. mars 2025].
- [111] A. Boyko, "BIM History Map", 2024. [Online]. Tilgjengelig: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "Definitions of BIM by Organisations and Standards", 27. desember 2023. [Online]. Tilgjengelig: <https://encyclopedia.pub/entry/53149>. [Dato for adresse: 5. mars 2025].
- [113] CAD Vendor, "OPEN BIM-programmet", 2012. [Online]. Available: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Referansedato: 30. mars 2025].
- [114] Wikipedia, "Industry Foundation Classes". [Online]. Tilgjengelig: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Dato for tilgang: 15. mars 2025].
- [115] Wikipedia, "IGES". [Online]. Tilgjengelig: <https://en.wikipedia.org/wiki/IGES>. [Dato for tilgang: 30. mars 2025].
- [116] A. Boyko, "History of CAD (BIM)," 15. desember 2021. [Online]. Tilgjengelig: https://miro.com/app/board/o9J_IaML2cs=/. [Dato for adresse: 24. februar 2025].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I," 8 Jan. 2021. [Online]. Tilgjengelig: <https://arxiv.org/pdf/2007.10951.pdf>. [Dato for tilgang: 5. mars 2025].
- [118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM," 13. april 2021. [Online]. Tilgjengelig: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be>. [Dato for adresse: 5. mars 2025].
- [119] D. Ares, "RETS i eiendomsbransjen: Hvorfor det er avgjørende for effektivitet og vekst", 17. desember 2024. [Online]. Tilgjengelig: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Dato for tilgang: 5. mars 2025].

- [120] "Flex token cost", 2024. [Online]. Tilgjengelig: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "Glem BIM og demokratiser tilgangen til data (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Tilgjengelig: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagment-und-Bay-Bauindustrie.pdf>.
- [122] D. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh og F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Tilgjengelig: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle>.
- [123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle," 1. januar 2018. [Online]. Tilgjengelig: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Dato for tilgang: 2. mars 2025].
- [124] Apple.com, "Pixar, Adobe, Apple og NVIDIA danner Alliance for OpenUSD for å fremme åpne standarder for 3D-innhold," 1. august 2023. [Online]. Tilgjengelig: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Besøkt: 2. mars 2025].
- [125] AECmag, "ADSKs granulære datastrategi", 25. juli 2024. [Online]. Tilgjengelig: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Dato for tilgang: 15. mars 2025].
- [126] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts", 24.11.2024. [Online]. Tilgjengelig: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a>. [Dato for adresse: 23. februar 2025].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Kampen om data og anvendelse av LLM og ChatGPT i bygg- og anleggsbransjen", 7. august 2024. [Online]. Tilgjengelig: ENG BIM Cluster 2024 | Kampen om data og anvendelse av LLM og ChatGPT i bygg- og anleggsbransjen. [Tilgang: 15. mars 2025].
- [128] "Jeffrey Zeldman Presents", 6. mai 2008. [Online]. Tilgjengelig: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Dato for tilgang: 15. mars 2025].
- [129] A. Boyko, "DWG Analyse med ChatGPT | DataDrivenConstruction," 5. mars 2024. [Online]. Tilgjengelig: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datatdrivenconstruction>. [Referansedato: 15. mars 2025].
- [130] McKinsey , "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [Online]. Tilgjengelig: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Dato for adresse: 30. mars 2025].

- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31. mars 2016. [Online]. Tilgjengelig: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Dato for tilgang: 15. mars 2025].
- [132] J. Bertin, "Grafikk og grafisk informasjonsbehandling", 8. september 2011. [Online]. Available: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Dato for tilgang: 15. mars 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking", [Online]. Tilgjengelig: <https://www.causeweb.org/cause/resources/library/r1266>. [Dato for tilgang: 15. mars 2025].
- [134] Ministrymagazine, "Hvordan vitenskapen oppdaget skapelsen", januar 1986. [Online]. Tilgjengelig: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Referansedato: 15. mars 2025].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23. mai 2017. [Online]. Tilgjengelig: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Besøkt 15. mai 2024].
- [136] "Slik bygger du en dataarkitektur for å drive innovasjon - i dag og i morgen", 3. juni 2020. [Online]. Tilgjengelig: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Dato for adresse: 15. mars 2025].
- [137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed-00011630>. [Dato for adresse: 15. mars 2025].
- [138] "Konverterere", 2024. [Online]. Tilgjengelig: <https://datadrivenconstruction.io/index.php/convertors/>.
- [139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1. januar 2017. [Online]. Tilgjengelig: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Dato for tilgang: 18. februar 2025].
- [140] "Rørledning i bygg- og anleggsbransjen", 2024. [Online]. Tilgjengelig: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedia, "Apache NiFi", 1. januar 2025. [Online]. Tilgjengelig: https://de.wikipedia.org/wiki/Apache_NiFi. [Dato for tilgang: 5. mars 2025].
- [142] n8n, "Gmail AI Auto-Responder: Create Draft Replies to incoming emails," 1. mai 2024. [Online]. Tilgjengelig: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Besøkt 15. mars 2025].
- [143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail", 1. mars 2025. [Online]. Tilgjengelig: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow>.

- api-google-sheets-and-gmail/. [Dato for tilgang: 15. mars 2025].
- [144] B. T. O'Neill, "Failure rates for analytics, AI, and big data projects = 85% - yikes!", 1. januar 2025. [Online]. Tilgjengelig: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.
- [146] T. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman", Quantitative Economics with Julia, 1. januar 2025. [Online]. Tilgjengelig: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Dato for tilgang: 1. mai 2024].
- [147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.
- [148] A. Boyko, "San Francisco. Bygg- og anleggssektoren 1980-2019", 2024. [Online]. Tilgjengelig: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects," 2024. [Online]. Tilgjengelig: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Preparing for the Future with Connected Construction", "Forbered deg på fremtiden med Connected Construction" [Online]. Tilgjengelig: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Besøkt 15. mars 2025].
- [151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22. mai 2017. [Online]. Tilgjengelig: <https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html>.
- [152] "Betingelser som kreves for bevaring av plantefossiler", 2024. [Online]. Tilgjengelig: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "BlackRocks Fink om obligasjoner, fusjoner og oppkjøp, den amerikanske resesjonen og valget: Fullt intervju," 2023. [Online]. Tilgjengelig: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 berømte AI-katastrofer", 2. oktober 2024. [Online]. Tilgjengelig: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Dato for tilgang: 15. mars 2025].
- [155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle", 1. januar 2025. [Online]. Tilgjengelig: <https://en.wikipedia.org/wiki/Kaggle>. [Dato for tilgang: 15. mars 2025].

- [157] Kaggle, "Titanic - Machine Learning from Disaster", 1. januar 2025. [Online]. Tilgjengelig: <https://www.kaggle.com/competitions/titanic/overview>. [Besøkt 10. mars 2025].
- [158] Ш. Johri, "Creating ChatGPT: From Data to Dialogue", 2024. [Online]. Tilgjengelig: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "Noen nyttige ting å vite om maskinlæring", 2024. [Online]. Tilgjengelig: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Siterbart sitat", [Online]. Available: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Dato for tilgang: 17. mars 2025].
- [161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models", 19. desember 2024. [Online]. Tilgjengelig: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Besøkt 25. mars 2025].
- [162] "NVIDIA kunngjør en større utgivelse av Cosmos World Foundation-modeller og fysiske AI-dataverktøy", 18. mars 2025. [Online]. Tilgjengelig: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Dato for tilgang: 25. mars 2025].
- [163] NVIDIA, "NVIDIA Isaac Sim," [På Internett]. Tilgjengelig: <https://developer.nvidia.com/isaac/sim>. [Besøkt 25. mars 2025].
- [164] M. Quarterly, "Hvorfor digitale strategier mislykkes", 25. januar 2018. [Online]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Dato for tilgang: 15. mars 2025].
- [165] M. J. Perry, "My Favorite Milton Friedman Quotes", 17. november 2006. [Online]. Tilgjengelig: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Dato for tilgang: 1. mars 2025].
- [166] J. A. Wheeler, "Information, physics, quantum: the search for links", 1990.
- [169] A. Boyko, "Lobbykrig og BIM-utvikling. Del 5: BlackRock er mesteren over alle teknologier. Hvordan selskaper kontrollerer åpen kildekode", 2024. [Online]. Tilgjengelig: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen og J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced Engineering Informatics*, 2024.
- [171] "Antall bedrifter i bygg- og anleggssektoren i Storbritannia i 2021, etter bedriftsstørrelse", 2024. [Online]. Tilgjengelig: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.

- [172] "5000 IFC&RVT-prosjekter", 2024. [Online]. Tilgjengelig:
<https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-dataldrivenconstruction-io>.
- [173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [Online]. Tilgjengelig:
<https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Referansedato: februar 2025].
- [174] *Lobbykrig om data i byggebransjen / Teknofeudalisme og historien om BIMs skjulte fortid*. [Film]. Tyskland: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT MED REVIT OG IFC | Automatisk gjenfinning av dokumenter og data fra prosjekter," 16. november 2023. [Online]. Tilgjengelig:
https://www.youtube.com/watch?v=ASXolti_YPs&t. [Dato for tilgang: 2. mars 2025].
- [176] M. & Company, "Tre nye mandater for å hente ut den fulle verdien av en digital transformasjon", 22. januar 2022. [Online]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Dato for tilgang: 15. februar 2025].
- [177] KPMG, "Construction in a Digital World", 1. mai 2021. [Online]. Tilgjengelig:
<https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-no.pdf>. [Dato for tilgang: 5. april 2024].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17. mars 2023. [Online]. Available:
<https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Dato for tilgang: 15. februar 2025].
- [179] O. Business, "Satya Nadella avslører 'Hvordan AI-agenter vil forstyrre SaaS-modeller'", 10. januar 2025. [Online]. Tilgjengelig: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Dato for tilgang: 15. mars 2025].
- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", 23. mars 2016. [Online]. Tilgjengelig:
<https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Dato for tilgang: 15. mars 2025].
- [181] UK Foreign, Commonwealth and Development Office, "Digital development strategy 2024 to 2030", 18. mars 2024. [Online]. Available: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Dato for tilgang: 15. mars 2025].
- [182] "Visjon og strategi i byggdesignbransjen", 7. november 2003. [Online]. Tilgjengelig:
<https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Dato for adresse: 5. mars 2025].

- [183] M. Bocharov, "Informasjonsmodellering", mars 2025. [Online]. Tilgjengelig: <https://www.litres.ru/book/mihail-evgenievich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [Dato for adresse: 15. mars 2025].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale", 2000. [Online]. Tilgjengelig: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Referansedato: 25. mars 2025].
- [185] CAD Vendor, "Open BIM Programme is a marketing campaign", 12. mars 2012. [Online]. Available: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Dato for adresse: 30. mars 2025].

INDEX FOR EMNE

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287

4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477

AIA, 289

AIM, 3, 289

AMS, 14, 84, 151, 153

Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494

Apache ORC, 62, 330, 378

Apache Parkett, 62, 67, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

B

BDS, 258, 259

Big Data, 9, 67, 245, 383

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

Avgrensningsboks, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

CAFIM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421

Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

CO₂, 72

CO₂, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

CoE, 56, 168, 169, 170, 477

Annenpilot, 114, 116, 459

CPIXML, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473

CQMS, 14, 84, 177, 178, 462

CRM, 109, 369, 459.

CRUD, 51, 459

CSG, 263

CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 356, 373, 377,

378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414
 DAG, 362, 363, 365, 366
 Dash, 320, 336, 337.
 Datastryring, 373, 395, 396, 398, 400, 401
 Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
 Data Lakehouse, 67, 373, 386, 387, 388
 Dataminimalisme, 373, 395, 396, 400, 401
 Data Swamp, 373, 395, 397, 401
 Datavarehus, 382, 383, 400
 Data-som-en-tjeneste, 487
 datadrevet, 50, 170, 460, 461, 484, 486
 DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
 DataOps, 170, 373, 398, 399, 400, 401, 475
 DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
 DGN, 8, 140, 186, 227, 357
 DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
 DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
 DXF, 8, 73, 277

E

ECM, 58, 175
 ECS, 142
 EIR, 289
 eLOD, 289
 ELT, 384, 385
 EPM, 14, 166, 198
 ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
 ESG, 196, 235, 236, 238
 ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
 Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
 Utdrag, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Fjær, 62, 123, 330, 378
 FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
 GIS, 58
 GLTF, 143, 278
 Google Regneark, 368, 370
 Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
 HiPPO, 29, 37, 95, 424, 477, 484, 490
 HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
 IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497
 IGES, 262, 263, 276.
 iLOD, 289
 IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484
 ISO 19650, 388

J

JavaScript, 320, 369, 378
 JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
 Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
 k-NN, 392, 393, 442, 445, 446, 447, 448
 KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
 LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494
 LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494
 Last, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475
 LOD, 287, 289
 LOI, 287
 LOMD, 287
 Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
 MCAD, 77, 78, 257, 284
 MEP, 14, 175, 492
 NETT, 234, 283, 284, 285, 296, 462
 Microsoft SQL, 65
 Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
 MRP, 11, 12
 MS Project, 70
 MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372
 NLP, 69, 394
 Ingen kode, 368, 369
 NURBS, 142, 282, 283, 284, 285
 NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285
 OCCT, 273
 OCR, 69, 128, 132, 134
 OmniClass, 154, 155, 156.
 åpen BIM, 142, 216, 219, 256, 261, 278, 291
 Åpen kildekode, 45, 55, 97, 98, 107, 108, 273, 275, 494
 UGLE, 267, 268, 269.

P

Pandaer, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494
 Parkett, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480
 PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
 PDM, 16
 PHP, 63, 339
 Rørledning, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476
 PLM, 16, 246
 PLN, 8, 272, 296
 Plotly, 320, 336, 337, 372
 PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487
 PMS, 84, 151, 233
 PostgreSQL, 63, 64, 65, 329, 395
 Power BI, 320, 372, 482
 private equity, 464, 487
 Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
 QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
 RDBMS, 63, 64, 65, 82, 89.
 RDF, 267, 268, 269
 RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
 RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
 ROI, 311, 317, 319, 321, 351, 370, 389, 479
 TURTALL, 14, 84, 151, 331, 333, 334
 RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
OMFANG, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Seaborn, 123, 320, 336, 337, 408, 412
SPARQL, 269
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363
TRINN, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transform, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

U

Uniclass, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474